

Does deregulation work?

European airlines are sheltering behind cosy cartels by saying that deregulation in the United States has merely restored the status quo. They should look again.

CURIOUS things have happened in the United States in the four years following the deregulation of airlines. At the outset, there were a dozen or so large and a host of smaller airlines competing for the domestic market within tight rules set by the regulatory authorities, whose effects were to say which carriers might fly how often between which airports. In the belief that regulations are irksome, and that their abolition would contribute to the more efficient use of a modern technology, the US administration decided to sweep away all but the rules requiring that airlines should comply with the regulations on which people's safety depends.

The outcome has had much of the excitement of a poker game. First, there sprang up a clutch of new airlines to service previously neglected markets. Frontier Airlines, for example, set out to serve the northern plains states and parts of Canada, and New York Air to demonstrate what the established airlines might have worked out for themselves, that there is a demand for air travel over short-haul routes along the eastern seaboard of the United States. Other new airlines set out to compete with the establishment over longer profitable routes: People Express, for example, first became a legend because of the cheapness of its transatlantic fares and then because of its style, that of a high-tech commune. But then, as the financial strains of competition began to tell, some airlines began to swallow others. The biggest player has been the previously regional airline called Texas Air, which has successively swallowed up Continental, New York Air, Eastern and, now, People Express (which had itself swallowed Frontier Airlines). So has the wheel turned full circle, with the United States still served by a dozen or so large and a host of smaller airlines?

That is what the cynics, disbelievers and opponents of airline deregulation say. But there is a crucial difference between the former and the present states of the US airline industry. Now, in particular, there is no reason why new airlines should not again spring up to serve parts of the US market that are either neglected or so profitable that newcomers can persuade their bankers that they have a sporting chance of winning profitable business. For as long as deregulation persists, this facility will be a thorn in the flesh of the complacent and the inefficient.

Sadly, the chance now seems to be evaporating that this simple truth would be appreciated in Europe, where air transport is regulated by means of bilateral cartels between governments on behalf of national carriers, many of them nationalized industries. The most obvious consequence of these arrangements is that air fares in Europe are scandalously high. There is no simple way of telling how much inefficiency shelters behind the network of arrangements requiring that pairs of airlines should share revenues on some profitable routes, that the numbers of airline seats on offer should be kept within predetermined limits or that, in at least one notorious case involving the Irish airline Aer Lingus, that there should be an annual payment to another airline for staying away from Dublin. Nor is it known how many potential travellers are kept on the ground by needlessly high fares. But the over-regulation of Europe's airline business must be one of the most effective ways of wasting European resources.

The European Commission has at last woken up to the dangers. Earlier this year, the Commission warned many of the most wayward airlines that their practices offend against the Treaty of Rome, the basis of the European Economic Community which is also a tract against restraints on competition. Before that, to their credit, the governments of Britain and of the Netherlands had taken the lead in trying to beat the system as it has ossified, by using cheap fares between London and Amsterdam as ways of undercutting other airlines' prices. Both governments have also done stalwart propaganda against the present network of corrupt arrangements, which is why many well-wishers had been hoping that the British presidency of the European Communities during the six months ending on 31 December would see a shake-up in the European airline business. Naturally, the problem cannot be solved simply. Established airlines are powerful vested interests in themselves, but they often command the affection of those who see them as extensions of the national flag and are an important source of jobs in depressed economies. But over time, it should be possible for a group of nations such as the members of the European Communities, united as they are by their wish to prosper, to put an end to this wasteful duplication and feather-bedding.

To judge from the British proposals for reform presented at a Community meeting last week, the well-wishers will be disappointed. The proposals consist largely of suggestions that airline routes now regulated by 50:50 revenue sharing should allow for some slack at the margins, so that ratios of 55:45 either way might be allowable. The steel seems to have gone out of those who were demanding, only a few months ago, that airlines with spare capacity should be free to deploy it where they chose. At least a part of the reason for the change of heart seems to have been the spectacle of the poker game played in the United States, but there are also dark hints that the British government is anxious not to change the rules in a way that would depress the planned sale of shares in the British nationalized airline, British Airways, later in the year. That would be so discreditable that it cannot be believed. □

Japanese distinction

Mr Nakasone can have meant only that Japanese are better educated than Americans.

Mr Yasuhiro Nakasone, the Japanese prime minister, is in many people's view the best thing to have happened to Japan since the invention of the transistor (at the then Bell Research Laboratories). His qualities as a modern Japanese (he was the minister responsible for science and technology in the crucial period fifteen years ago) and as a courageous politician have won him, in the past few weeks, his party's consent that he should stay on as leader (and thus as prime minister) for an extra year. How does a sensible fellow like that become embroiled in a peculiarly Japanese version of the old nature/nurture argument about the evolution of human intelligence?

With all the fuss of the past few days, it is probably now too late to tell precisely what Mr Nakasone told the young members



of his political party at their weekend meeting on the slopes of Mount Fuji two weeks ago. It seems, however, that he said (in English) that Japan is an "intelligent society" and that the United States, in some sense of the phrase, is less so on account of the "minorities" it includes. Mr Nakasone has apologized to anybody who have thought he meant to say that minorities in the United States, black or otherwise, are in some sense inferior to Japanese. Of necessity, he has not withdrawn the nub of his assertion that there is something special about Japan. This truth, of course, is widely recognized as such; even if the pace of economic growth has slackened in the past few years, Japan's economic transformation in the past 40 years counts as a real-world miracle. But is it also possible that Mr Nakasone has been unduly influenced by the report by R. Lynn four years ago (*Nature* 297, 222; 1982), suggesting that the improvement (increase) of the IQ of children in Japan in the years since the Second World War had been more rapid than that among schoolchildren in the United States during the same period? Lynn's study was widely reported in Japan when it appeared.

Next to the rapid economic growth of recent decades, almost the most remarkable feature of Japanese society is its self-conscious regard for its standing in the eyes of others. No doubt this curiosity stems in part from the general knowledge that almost within living memory, Japan was a closed society, with hardly any links with the rest of mankind, which accounts for its racial and cultural homogeneity (some vociferous Koreans notwithstanding).

Prudent commentators also note three features of modern Japan of which Mr Nakasone is well aware. First, people's devotion to the education of their children is literally unparalleled; the participation in higher education is no less than 36 per cent. Second, Japan has not spent much time or resources on the development of social services, with the consequence that the disadvantaged are literally so, while those still at work save all they can against the time when the calamity of retirement strikes them (which is why capital investment is so high). Then, third, Japan is in many ways an inefficient economic system, with a distribution system employing many more people than would be needed in most other parts of the world (which may compensate for the lack of social security), urban roads perpetually clogged with traffic and urban trains perpetually crammed with people travelling to and from their work.

What Mr Nakasone should have said, and what it must be presumed he meant to say, is that the most distinctive feature of modern Japan that distinguishes it from the United States is the zeal with which education is pursued. Because, these days, nobody supposes that measured IQ is an unchangeable attribute of individuals, Japanese education may, by itself, account for the findings reported by Lynn four years ago. But that does not imply that Japan's advantage could be easily repeated elsewhere; cost apart, this generation's education is also an investment in its successor. Japan's homogeneity, cultural and ethnic, may be less of an advantage than the common denominator of Nakasone's remarks implied; does anybody at this stage doubt the effectiveness of the melting pot of the United States in the later nineteenth century as an engine of economic and social development? The only simple ways of comparing complicated societies are by definition simplistic. Maybe Mr Nakasone meant to say that as well? □

Death of a submarine

The loss of a Soviet nuclear submarine is a reminder of the seriousness of the world we live in.

THE good news first. Just two weeks ago, most governments with pretensions to nuclear power, military or otherwise, put their names to two conventions negotiated during August, in the aftermath of the Chernobyl accident, which among other things require that those responsible for nuclear accidents should

promptly let others know. The language of the conventions is such that military accidents may or may not be covered; much depends on how those responsible decide to interpret the documents they have signed in the light of the circumstances in which they find themselves. On this latest occasion, the Soviet Union has done everything that might be expected of it. The United States was told there had been an accident on board a nuclear submarine a few hours after it had happened. The Soviet authorities have also said that the reactor that had driven the vessel had been shut down before the ship was scuttled. The nuclear weapons on the vessel will finish up on the deep ocean bed, where their load of fissile material will be released in the course of the next few years, thus providing an unwitting test of claims that sea-bed disposal is safe. (Two other Soviet submarines have been lost on earlier occasions, but so far as can be told they have not yielded detectable contamination.) The International Atomic Energy Agency, to which Chernobyl gave a new lease of life, can hardly have hoped there would have been such a quick and successful test of people's willingness to adhere to its conventions on nuclear accidents that have international implications.

The bad news (but that is also good news in one sense) is that a Soviet submarine built for attack should have run into trouble a few hundred miles from the eastern seaboard of the United States. That is a reminder, also timely, of the role that nuclear deterrence plays in the present strategic balance. A little thought, of course, will show that the western mid-Atlantic is one of the places in the oceans at which Soviet submarines would be stationed. There will be others off the Pacific coast, just as there will be US (and perhaps even French and British) nuclear submarines in the western Pacific and the North Atlantic. Submarines, of course, are less conspicuous than land-based strategic missiles, which have in any case attracted most attention in the past decade because of the concentration of novel types in Europe. But the secretiveness of submarines is also their strategic strength, recognized by all sides, which is why their abolition is not on the agenda at this weekend's hasty meeting in Reykjavik. During the negotiation of the present Strategic Arms Limitation Treaty (SALT II), the United States was even pressing for a greater reliance on submarines. Whatever happens next weekend, they will continue to lurk off people's vulnerable coastlines at least until they can be spotted there more easily than at present. (Some people would give that ten years, but realists see more future in submarines.)

This has a bearing on the debates in the past few weeks on the role of nuclear weapons in the defence of Europe. Even though the next general election may be as much as eighteen months away, British voters have on offer a spectrum of policies from which to choose. The government is at one end and the Liberal Party (inconveniently half of the association with the Social Democratic Party (SDP) called the Alliance) at the other. The Labour Party would get rid of other people's nuclear weapons, and maybe even those it would inherit if elected; the SDP would keep nuclear weapons as such, but not buy the new Trident nuclear submarines, seeking some nuclear alliance with France instead. For good measure, the Labour Party would also "phase out" civil nuclear power stations, most probably by letting those now in service come to the ends of their working lives.

In a curious way, all these positions are given the character of charades by the sinking of the Soviet craft north of Bermuda. Nobody supposes that the loss of a single nuclear vessel will take the edge off Soviet strategic power, but such an event would quickly make the threat posed by the smaller forces seem the opposite of what it is meant to be — credible. Both Britain and France could make a lot of mischief with their existing nuclear forces if the big boys allowed them to do so. But because the chance of that happening is small, even declarations that unilateral disarmament will provide an example for others seem puny gestures. Should not these smaller nuclear powers at least decide more clearly than they have done so far what these dangerous weapons are for? []

US semiconductors

Worries about agreement on trade with Japan

Washington and Tokyo

JUST two months after the United States and Japan signed a comprehensive agreement governing trade in semiconductors, the US semiconductor industry is accusing Japanese companies of violating the agreement by dumping memory chips on world markets below their "fair market value".

The agreement, reached after months of negotiations, requires Japan to prevent its industry from selling below cost, or dumping, erasable programmable read-only memories (EPROMs) and dynamic random-access memories (DRAMs) on world markets. Japan also agreed to help US companies to achieve more than the 10 per cent share of the Japanese market they now enjoy.

In exchange, the United States agreed to drop two anti-dumping suits that would have levied large import duties on Japanese EPROMs and DRAMs. The United States also ended an investigation into unfair trading practices by Japan that could have resulted in trade retaliation.

The US industry is satisfied that Japanese chips are no longer being sold on the US market at unfair prices. But it says dumping continues in other international

markets. The industry claims to have proof of Japanese chips being dumped in Hong Kong; it is also upset by chip prices in Japan itself, where 256k DRAMs can be purchased for as little as ¥280 (\$1.80), below the figures for fair market value established when the agreement was signed.

Although the agreement does not extend to Japan's domestic market, US officials say they might take action if the low prices prevent growth of US companies' market share in violation of the agreement. The irony of the United States trying to regulate Japanese domestic prices has not been lost on US trade officials.

When the agreement became effective last August, the US Department of Commerce estimated that some Japanese companies were selling their chips in the United States at less than half their cost, so that prices were expected to rise after the agreement took effect. They did, but not dramatically. Dataquest Inc., a semiconductor information service, says 256k DRAMs cost \$2.40 before 1 August, but rose only to \$2.85 afterwards. It is now possible to buy 256k DRAMs from a US retail distributor for only \$2.75.

Prices in Japan, in contrast, plummeted by nearly 11 per cent in the past month according to a report in Japan's leading economic newspaper, the *Nihon Keizai Shimbun*, and further falls are expected due to decreased demand, users' demands for lower prices and a surplus of chips created by the agreement — trading companies saddled with chips they could not sell because of increased export prices have unloaded their excess inventory on the domestic market driving down prices to as low as ¥250 (\$1.80), far below the "fair market values" of \$2.50–7.00 stipulated in the agreement.

At the end of this week, the Commerce Department will release to Japanese companies revised figures for the fair market value of the chips covered by the agreement. Those new figures will be effective until the end of 1986, and are expected to be lower than the estimates originally used to conclude the trade agreement. For example, fair market values for 256k DRAMs were expected to range among manufacturers from \$2.00 to \$4.00.

Commerce deputy under-secretary Gilbert Kaplan nevertheless disputes the industry's claim that there are problems with the agreement. Initial enquiries about sales to countries other than the United States were answered "very satisfactorily", he says. The first of the regular consultations called for in the agreement could take place as soon as the end of this month. Kaplan insists that the emergency consultations sought by the US industry are unnecessary.

The agreement was given a vote of confidence last week, when the Idaho-based Micron Technologies dropped a large anti-dumping lawsuit against six Japanese manufacturers. Joe Parkinson, president of Micron Technologies, says his company dropped its suit because he believes the agreement will work.

Not all elements of the US industry are enthusiastic about the semiconductor agreement. The Computer and Business Equipment Manufacturers Association, whose members are major users of memory chips, fears that large price increases will result from the agreement. The association is also concerned that the Commerce Department will be hard pressed to gather all the data necessary to prepare accurate estimates of the fair market value of Japanese chips. Meanwhile, the main beneficiaries of the agreement appear to be Japanese computer manufacturers who are enjoying rock-bottom prices in Japan. Ken Flamm of the Brookings Institution in Washington believes that if the Commerce Department's calculations of fair market values are not close to the mark, there will inevitably be problems. He says that if regulated price differential gets too large, "that's when smuggling develops".

Joseph Palca & David Swinbanks

Sino-British collaboration in Tibet



LAST year, the British Royal Society and Chinese Academia Sinica jointly supported an expedition to the Tibetan plateau, and the preliminary results are published today (see p. 501). The low hills above (altitude 5,050 m) are composed of red sandstones that were deposited on the Qiangtang Terrane, a continental fragment that collided with the southern Eurasian continental margin about 200 million years ago. The Chinese-British expedition identified three such fragments

along their traverse, which followed the 1,150-km-long Lhasa-Golmud highway. The sandstones shown here were deposited about 50 million years ago, as a result of intracontinental shortening during the collision of India with Tibet. One of the major results of the expedition is that the thickening of the Tibetan plateau crust can be explained by such internal shortening, and does not require extensive underthrusting of the plateau by Indian continental crust. □

Soviet Union

Freedom at last for Orlov

Dr Yuri Orlov, physicist, human rights activist and ex-Soviet citizen, arrived in the United States last Sunday after serving nine years of a 12-year sentence (seven years prison plus five years exile) on charges of slandering the state. Since his arrest in 1977, Dr Orlov had been the centre of campaigns by fellow physicists ranging from a "formal counter-trial" staged at the Institute of Physics in London to "Free Orlov" T-shirts produced by sympathizers at the CERN laboratory near Geneva and worn conspicuously whenever Soviet visitors were expected.

Orlov, who was born in 1923, served in the Red Army during the Second World War and afterwards entered Moscow University, graduating in physics in 1951. He then began graduate research at the Institute of Theoretical and Experimental Physics of the Soviet Academy of Science. In 1955, however, after Khrushchev's "secret speech" to the Twentieth Party Congress denouncing Stalinism, Orlov thought that the way was open for democratization of the party (he had become a member while at university). But his enthusiasm for democracy was misplaced — he lost his job and was expelled from the party. Eventually, he was employed in Armenia, where he took his Candidate's degree (PhD), working on particle accelerators.

In 1972, he returned to Moscow, where, in touch with Dr Andrei Sakharov, his enthusiasm for human rights and democracy gained new force. Within six months he was again dismissed, and, although not a Jew, as a result became friendly with the Moscow "refusniks" expelled from their professional posts after filing applications to emigrate to Israel. For a time he took an active part in the "Sunday seminars" founded by Alexander Voronel to help the refusnik scientists keep up some kind of intellectual life.

Following the signing of the Helsinki Final Act on 1 August 1985, Orlov and a group of friends established the Moscow "Helsinki Monitoring Group" to observe and report on Soviet fulfilment of the pledges on individual freedom and East-West contacts. This initiative, which by Soviet norms was outside the law, was soon copied by a Ukrainian group in Kiev, and later in Poland.

Orlov's group produced 18 reports on Soviet breaches of the Accords, but its members were one by one picked up by the police and put on trial. Orlov was arrested in February 1977 and tried in May 1978 under Article 190/A of the Russian penal code ("anti-Soviet agitation").

Throughout his imprisonment and exile, Orlov's name was kept in the public eye by supporters and sympathizers, not

all of whom acted with due forethought. In the United States, his former colleague in the human rights movement, Valentin Turchin, called for a boycott of the 1980 Olympics unless Orlov was released, although the connection between Orlov and athletics was not immediately apparent, and even some scientists who decided, on Orlov's behalf, to cut their professional contacts with the Soviet Union felt that it was not fair to force athletes to sacrifice a once-in-a-lifetime chance.

At about the same time, well-meaning Russian emigrés who heard that Orlov was keeping up his theoretical work in prison and had managed to send some of his notes abroad (a facsimile of some of his calculations appeared in *Nature*) claimed that this was, in fact, not mathematics but code, with the not surprising consequence that Orlov was thereafter forbidden to do any science.

And when his seven-year prison sentence expired, there were ingenuous suggestions that he should be allowed to serve

his five years exile not in Siberia but abroad, at CERN, a proposal that must surely have amused the Soviet authorities, as such assignments are among the most sought-after rewards in Soviet science.

In the end, of course, Orlov's release came in the most face-saving manner open to the Soviet authorities — as part of an exchange of undesirables, intended to clear the air for the Reagan/Gorbachev not-quite-summit in Reykjavik. His future, for the moment, remains unclear. Doubtless, like his colleague from the Helsinki Committee, Anatolii (now Nathan) Shcharanskii, he will be expected to make a triumphal tour of the committees, in Switzerland, Canada, the United Kingdom and the United States, that have worked so vigorously on his behalf. But, in spite of his poor health and the undoubted hardships of the past nine years, Yuri Orlov is, by the standards of Soviet science, by no means at the end of his creative life. Several universities and learned institutions have already indicated that they would welcome him, if necessary in a visiting or part-time capacity, should he wish to resume his scientific career.

Vera Rich

US embargo

British supercomputer deadlock

OFFICIALS at the British Department of Trade and Industry will make fresh attempts this week to arrange a meeting with their counterparts from the US Department of Commerce to secure an export licence for a supercomputer destined to be the mainstay of advanced academic computer research in the United Kingdom. The difficulty is that the United States, fearing that advanced technology might fall into the hands of alien nations, principally the Soviet Union, is seeking to impose strict constraints on access.

British authorities consider the US approach to be heavy-handed, as the United Kingdom is a member of COCOM, the Coordinating Committee for Multilateral Export Controls, composed of the North Atlantic Treaty Organisation countries, except Iceland and Spain, in partnership with Japan. The committee controls the export of strategically sensitive high-technology products and knowledge to the Warsaw Pact countries and China through the issue of export licences. No more controls are needed, say the British.

The research supercomputer, a Cray X-MP system which, with ancillary electronics, will cost about £48 million over the next five years, is being purchased on the recommendation of a study commissioned by the Advisory Board for the Research Councils, the University Grants Committee and the Computer Board for Universities and Research Councils. The study, headed by Professor A. Forty of the Uni-

versity of Warwick, was published in June last year and recommended the joint purchase and use of a supercomputer for British university research.

But the project has reopened many old wounds and raised, once more, the spectre of US "extraterritoriality". The US Department of Commerce has proved inflexible, the British believe, in its unnecessarily tight controls on high-technology exports and in its insistence that resale of high-technology equipment needs supplementary licences.

The British have never agreed. Trade officials in the past three years have been locked in battle over the issue with the United States. The situation appears to be getting worse. The US authorities are determined that the controls will not be relaxed but intensified.

Since the spring, the United States has intensified its efforts in offshore control, especially by its requirement that some British companies should be "inspected" by officials from the US Department of Commerce to confirm their "legitimate" use of licensed US technology. Last spring, trade minister Alan Clark acknowledged that the issue had become a dispute between the two governments, and said that the British government was still then considering whether to agree to the proposed inspections. That issue still remains outstanding and may well be a bargaining counter at the supercomputer negotiations.

Bill Johnstone

Europe

Research budgets threatened

POSSIBLY over-ambitious British attempts to persuade Brussels to streamline its research programme, a general reluctance to increase budgets in the European Economic Community, and political brinkmanship in Brussels are all conspiring to threaten a decrease in European research spending in 1987 — rather than the near-doubling requested by Community officials earlier this year.

This is the spectre that is beginning to haunt the Brussels research directorate just eight weeks before European ministers are due to decide on a new "framework programme" for research and development, a programme for which Brussels has requested 7,735 million European Currency Units (ECU; £5,500 million) plus a contingency fund amounting to another 15 per cent, against current spending of over 4,000 million ECU.

The immediate problem is that many existing programmes are coming to the end of their current term. Research to help developing countries, for example, ends in December, and so officially does the European stimulation programme on research in information technology, ESPRIT, although that programme has some unspent cash in its coffers. To deal with a possible hiatus if the framework programme is not agreed in December, Commission research officials had proposed some interim topping-up. But now this has been vetoed from the top by Commission President Jacques Delors.

At stake, according to Delors, is the whole principle of the framework programme, which is in turn embodied in the European Single Act, the reformed treaty governing Community affairs which is still awaiting ratification in most national parliaments. The Single Act for the first time makes science and technology a proper activity of the commission (previously it slid by under the "any other business" Article 235 of the Treaty of Rome), and it demands that ministers agree on an outline enabling programme and budget (the framework) before detailed elements of the programme are fixed. Although the existing framework programme is drafted so that it can go through under Article 235, Delors has told research commissioner Karl-Heinz Narjes that the principle of framework first, detailed decisions second must be followed.

In practice this is brinkmanship; for Mr Geoffrey Pattie, Britain's technology minister, who is currently president of the council of research ministers, has nailed his colours to the sorting out of Brussels research by December. Mrs Margaret Thatcher, the British prime minister, is due to speak on the achievements of Britain's six-month presidency of the Council

of Ministers a few days after the crucial research council meeting at which the framework programme is due to be decided: if it is not, given Delors' position, she will be describing a near-descent into anarchy.

But meanwhile, Pattie is trying desperately but bravely to get Brussels research plans into the lean, industrially oriented and objectively assessed shape he thinks fit, while at the same time trying to juggle 12 independent national demands about the content and scale of Brussels programmes. The main outstanding issues are, however, the level of funding and degree of assessment of research, on both of which Britain is hawkish, and Delors no doubt feels that a little extra pressure on Britain may result in a slight softening of British resolve, although West Germany and France would still be tough nuts to crack.

Britain may remain equally tough. Pattie outlined his view of Brussels coordinated research last week, in an ESPRIT review meeting; and, although he was speaking so close to the decisive council meeting, he remained ambitiously radical. ESPRIT is widely considered one of Brussels' great successes, in getting wary and nationalistic information-technology businesses involved in joint research. Many in the programme expect its expansion. Pattie dropped strong hints to the contrary. The £350-million Alvey programme in a similar field is also coming to an end. A couple of times Pattie indicated the advantages of this programme: it was more rigorously assessed by peer review, and national management could be more efficient. The extension of ESPRIT in its second phase to projects that were much closer to the market place, as the Commission had proposed, also "created real problems" for some member states. Rather than merely extend a successful programme, the Commission should just recognize its success and move the ESPRIT funds into other areas where a stimulus to cooperation would be useful. Here Pattie mentioned not research but investment and marketing. Any second phase of ESPRIT itself would have to be more concerned with small companies than the large ones in which most ESPRIT research is now carried out. Brussels, it seemed, should now steer away from the big boys who now had a real job to do.

European Commission ambitions may thus be curtailed, with West Germany and France following the British line. But there are another nine members of the European Community, more of whom support the Commission proposals than do not, and Delors' brinkmanship could be giving them weapons. **Robert Walgate**

Plagiarism

Earth-shaking controversy

Washington

CONTROVERSY is shaking the National Science Foundation (NSF)'s decision to put a \$25 million earthquake research centre in Buffalo, New York. The award, made in August to the State University of New York (SUNY) at Buffalo, has been sharply criticized by a California team that competed for it. Now a scientist at the California Institute of Technology says parts of SUNY's proposal were copied from his own work, and California senators recently directed the General Accounting Office (GAO) to audit the foundation's award procedure.

NSF's decision initially put the California team's nose out of joint because, along with a greater supply of raw material, California researchers also feel they have an advantage in expertise. (Since 1900, California has had nearly 300 times as many earthquakes measuring more than four on the Richter scale as has New York.) The principal investigator at the University of California, Berkeley, questioned NSF about its review process late in August. Several weeks later, Paul Jennings, a Caltech professor who co-authored the Berkeley proposal, found sections of SUNY's proposal that repeated verbatim some of his contributions to a 1984 report, and he pointed this out to NSF and California senator Pete Wilson.

The foundation then asked SUNY to respond to Jennings' objections, but the university's retort has been muffled by the absence of the proposal's principal author, Robert Ketter. Meanwhile SUNY's provost has told NSF that the parts of the proposal that appear to be copied "constitute less than 2 per cent" of the total 53-page work and are not essential to the original science that won the grant. SUNY says that Jennings' work was credited, although not meticulously.

Even if all sides agree that the material was lifted from Jennings' writing, NSF may still be in a quandary as to what to do about it. Jennings thinks the proposal should be disqualified and the review process began again. A SUNY spokesman says the Californians merely have a case of "sour grapes". Wilson and New York Representative Jack Kemp have locked horns over the GAO audit. NSF is conducting an internal investigation to clarify its own review procedure, but has already sent more than \$800,000 to SUNY investigators.

And all are still awaiting the word from Ketter, who returned from abroad late last week to a situation as tumultuous as the earthquakes he planned to study.

Karen Wright

Genetic manipulation

Relaxed rules provoke anger

Washington

THE Recombinant DNA Advisory Committee (RAC) of the US National Institutes of Health last week voted to relax environmental release reviews for some recombinant organisms. The new guidelines exempt organisms with deletions and certain rearrangements in genetic material from RAC evaluation before field testing. In practice the change will affect few investigations, but its principle might have greater, and more unpredictable, ramifications.

RAC stopped reviewing laboratory research with such engineered microbes years ago, and now claims there is no scientific basis for discriminating against the deliberate release of recombinant organisms whose minor genetic alterations could arise through natural processes. The committee determined that deletions and rearrangements of non-chromosomal or viral DNA would not pose a risk any greater than that introduced by recombination in nature. Investigators wanting to release such organisms into the environment need no longer notify RAC of their intentions. But most experiments will still require the scrutiny of other regulatory agencies.

But some critics are calling this stance "a cop-out". Two RAC members voted against the proposal on the grounds that even deletions and rearrangements harbour unknown risks and could, theoretically, have unforeseen and undesirable consequences. One of the dissenters, ecologist Fran Sharples of the Oak Ridge National Laboratory, says the committee's mandate is to judge recombinant processes rather than on the nature of the products. She fears that "there is now a group of field-release experiments that won't be reviewed by anybody".

The National Cancer Institute biologist who proposed the amendment disagrees. "In and of itself, it won't affect that many people", says Susan Gottesman. "It's more a statement of principle than anything else". Gottesman is dismayed at the rocky road the recombinant ice-minus bacteria and pseudorabies vaccine have had to travel to obtain permission for release. She hopes RAC's motion will persuade other agencies to ease their reviews.

Fred Rapp, an RAC member from the Pennsylvania State University, warns that loss of public confidence may be one of the most serious problems the exemptions will engender, because subtle but important restrictions to the exemptions may not be brought before the public eye.

Indeed, Gottesman's strategy could backfire if regulatory agencies try to stiffen their requirements to take up the slack, though sparse, left by the RAC decision. The committee has already ruffled

some feathers at the Environmental Protection Agency, which asked RAC to hold the vote on the proposal until its own biotechnology committee was established sometime this fall. RAC did not.

Congress may likewise take a knee-jerk dislike to RAC's action. A report due this week from the House of Representatives science and technology subcommittee will probably call for greater control, based on four hearings on biotechnology regulation held between December 1985 and last

July. A staff spokesperson says Congress does not doubt RAC's scientific expertise. But "RAC doesn't do things for scientific reasons, it does things for political reasons. And this time it made a mistake".

RAC's resolution still needs the signature of the institutes' director James Wyngaarden to be finalized. Meanwhile, RAC members are also debating a redefinition of "recombinant" organisms in general that would exclude most cases in which foreign DNA was not introduced into the host genome. So while RAC seems bent on loosening its regulatory reins, it is not clear that other agencies will follow suit.

Karen Wright

Netherlands research

EMBL membership reviewed

Waalre, the Netherlands

CONTINUATION of Dutch membership in the European Molecular Biological Laboratory (EMBL) in Heidelberg seems to be in doubt. Science Minister Willem Deetman has asked the Royal Academy of Sciences and the Advisory Council for Science Policy for a "solid evaluation" of Dutch membership and a review committee has been set up, composed of members of both scientific bodies. It will be seeking to answer a number of questions concerning the way that EMBL works at present, how it might be made to work better and

seen with some reservation". The council was also critical about the European Communities' Joint Centre for Research (in Petten, the Netherlands), saying that its work should be evaluated. The council acknowledges the international character of science, but also notes the disadvantages inherent in bureaucratic institutions. The Netherlands gives 5 per cent of its research and development budget to international organizations. The European Space Agency and CERN get most of the annual DFL 200 million. The council warns that the Dutch research and development effort is lagging behind that of other industrialized countries; funding is 2 per cent of gross national product as against 2.5 per cent, or more elsewhere.

That warning is echoed by the findings of a report from examiners from the Organisation for Economic Cooperation and Development which also advised the government that it should increase its commitment to the country's research base (*Nature* 320, 673; 1986). "The industrial structure of the country and the very big expenditure of five large firms do not suggest any grounds for believing that the Netherlands economy can afford to rest on a plateau". However, on 6 September, the government, with a cabinet composed of the same political parties as in May (before the elections), proposed a budget for 1987 that allows only an extra DFL 55 million on the research and development budget (currently DFL 4,100 million). With the DFL 4,600 million from industry, the Netherlands will be spending a total of 2.2 per cent of its gross national product on research and development in 1987.

Deetman promises that he will promote the use of research facilities in other countries by Dutch scientists and encourage cooperation between European institutes. If international cooperation is to be fostered, there may yet be a chance that the Dutch will not lose heart over EMBL.

Casper Schuurings



Willem Deetman — re-thinking Dutch membership.

what the Netherlands gets out of it compared with what it might get from spending the money at home. Another question is whether EMBL needs an international review committee. The minister expects an answer before 15 May 1987.

The Advisory Council for Science Policy emphasized in its yearly advice to the government in June that continued financial participation in EMBL "should be

Occupational radiation

British nuclear workers cleared

AN independent study of mortality among workers at the British nuclear reprocessing plant at Sellafield (previously known as Windscale) has confirmed earlier claims by the operating authorities that the workforce at the plant is even less susceptible to diseases believed caused by radiation than the general population of Britain. But the study, carried out by Dr Peter Smith and Ms Alison Douglas of the London School of Hygiene and Tropical Medicine, suggests that myeloma and, unexpectedly, bladder cancer are significantly more common among those exposed to large doses of radiation.

The study is founded on an attempt to trace all the 14,327 people who worked at the site between 1947 (when operations began there) and the end of 1975. Only four per cent of the workforce could not be traced. Of the remainder, 2,277 have died, a quarter from cancer of all kinds. The report, published in the current issue of the *British Medical Journal* (293, 845; 1986), says that in the period up to the end of 1983, mortality among the Sellafield workers was 2 per cent less than that in the general population of England and Wales and 9 per cent less than that in Cumbria, where Sellafield is situated.

For many years, British Nuclear Fuels, the company now operating the Sellafield plant, and its predecessors have pointed to the comparatively low mortality among Sellafield workers as evidence that radiation exposures there have not been overtly damaging. But the evidence has not carried the weight it might have done because it had been assembled by interested parties and because, as Smith and Douglas also acknowledge, selection of only healthy people for work at reprocessing plants must bias statistics of mortality.

One of the surprising features of the report is the evidence it provides that, despite the specialist character of the work at reprocessing plants, only just over a half of the workforce employed before 1976 had worked at the plant for five years or more.

Radiation records were apparently available for nearly three-quarters of the Sellafield workforce (the remainder being people exposed only infrequently to radiation). Figures for radiation doses due to internal radiation have not been used in the study. The average radiation dose for those for whom film-badge records are available is given as 124 mSv (12.4 mrem) accumulated until the end of 1983. The total collective radiation dose for the whole workforce is estimated at 1,260 Sv between 1947 and 1983. The report shows a steady reduction by about a half of the average annual radiation dose to workers at Sellafield since the early 1970s.

One of the findings of the study is that

Sellafield workers were less likely to die in the first five years of their employment, whether of cancer or of some other disease, than members of the general population, presumably one of the effects of the selection of healthy people for employment. According to the study, the lower incidence of conditions such as bronchitis among those exposed to the larger doses of radiation may result from similar selection effects, possibly because people with chest conditions are discouraged from becoming radiation workers or because their higher socio-economic status diminishes their use of cigarettes.

For cancer mortality, the report says that the risk of death from cancer among Sellafield workers was 5 per cent less than

in the general population of England and Wales and 3 per cent less than among the population of Cumbria.

The only statistically significant causes of death from cancer in respect of which the Sellafield workers differ from the general population of England and Wales are bladder cancer (14 cases), multiple myeloma (7 cases), leukaemia (10 cases) and all lymphatic and haematopoietic carcinomas (27 cases). Correlating radiation exposure with the occurrence of these conditions (after a time-lag of 15 years), the study arrives at numerical values for the supposed linear relationship between radiation exposure and mortality, but the significance of the coefficients is not so great that the association can be taken as more than cause for further investigation. Similar results for multiple myeloma, have been reported in a study of the workforce at Hanford, United States. □

Space shuttle

Programme for 1988 announced

Washington

MILITARY missions and communications satellites will be the predominant payloads when US shuttle flights resume in 1988 according to a manifest released last week by the National Aeronautics and Space Administration (NASA). But three of the first nine flights will be reserved for science missions, including the long awaited Hubble Space Telescope.

The first shuttle launch, now scheduled for 18 February 1988, will place a tracking and data-relay satellite (TDRS) into orbit. Once in place, these satellites will make up NASA's primary system for space communications. There is only one such satellite currently operational. Challenger was carrying a TDRS satellite when it exploded last January, and a third satellite will be launched in September 1988.

The Hubble Space Telescope will fly aboard the shuttle Atlantis on 17 November 1988, according to the new manifest. Robert Milkey, assistant director for programme management at the Space Telescope Science Institute (STSI) says that although only one operational TDRS satellite is essential for the space tele-

scope's functioning, two will be extremely useful. Milkey admits he would like to have an earlier launch date, but finds it "very gratifying" to be the fifth mission after flights resumed.

The Ultraviolet Astronomy Telescope (ASTRO-1) is scheduled for launch in January 1989, and the Venus radar mapper Magellan in April 1989. For Magellan, this means a longer type-4 trajectory, with an arrival at Venus in mid-1990. There is also a slot in November 1989 for a planetary mission. No decision has yet been made whether this will be given to Galileo, the Jupiter probe and orbiter, or Ulysses, the solar polar mission, but it is now clear that both will not be launched in 1989. NASA will face political pressure from the European Space Agency to launch Ulysses as soon as possible, but it also faces considerable cost in keeping Galileo on the ground. Although NASA administrator James Fletcher says that a decision on launch assignments for Galileo, Ulysses and the Mars Observer spacecraft will be made in 1987, for planning purposes a decision will have to come earlier than that.

Joseph Palca

Date	Payload
1988 February	TDRS-C, communications satellite
1988 May	Department of Defense
1988 July	Department of Defense
1988 September	TDRS-D, communications satellite
1988 November	Hubble Space Telescope
1989 January	ASTRO-1, ultraviolet astronomy telescope
1989 March	Department of Defense
1989 April	Magellan, Venus radar mapper
1989 June	Department of Defense
1989 June	2 global positioning satellites and a materials science laboratory
1989 July	Department of Defense
1989 August	Department of Defense
1989 August	2 global positioning satellites and a materials science laboratory
1989 November	Planetary mission
1989 December	Space life sciences laboratory

US-Soviet space

Towards renewed collaboration?

Washington

A SECRET meeting took place last month in Moscow between representatives of the US National Aeronautics and Space Administration (NASA) and the Soviet Institute for Space Research. According to a source close to the participants, the purpose of the visit was to lay the groundwork for the resuscitation of a bilateral agreement on cooperation in space between the two countries. The previous agreement lapsed in 1982.

The US delegation was headed by Lew Allen, director of the Jet Propulsion Laboratory in Pasadena. Joining him from NASA headquarters in Washington were deputy associate administrator Sam Keller, chief scientist Frank McDonald and Peter Smith, chief of the international programme policy office. Joe Kerwin from the Johnson Space Center in Houston was also there, as were representatives of the Departments of Defense and State and the National Security Council. The meeting was arranged by the Soviet Academy of Sciences, and representatives of all key space institutes were present.

The meeting's chief goal was to discuss

potential areas for cooperation, including planetary exploration, life sciences, Earth sciences, solar sciences and astrophysics. No startling new missions were formally considered, although there was informal discussion of a proposal to return a sample from the surface of Mars.

A formal treaty for cooperation in space exploration was signed by the United States and the Soviet Union in 1972. Among the most ambitious efforts carried out was the Apollo-Soyuz Test Project in 1975, when US and Soviet spacecraft docked in orbit and the two crews exchanged visits. Other cooperative efforts included US biomedicine experiments on Cosmos satellites and the sharing of data from Soviet Venera probes to Venus.

After the treaty's first five years, both countries agreed to a five-year extension. But, from 1977 to 1982, relations deteriorated because of the Soviet presence in Afghanistan, the US establishment of diplomatic ties with China and US charges that the Soviet Union had violated the biological weapons convention. The treaty was not renewed in 1982.

By the mid-1980s, the United States was again interested in cooperation with the Soviet Union. On 30 October 1985, President Reagan signed a resolution sponsored by Senator Spark Matsunaga (Democrat, Hawaii) that called for reestablishing the agreement and exploring new areas for "East-West ventures in space". Until recently the Soviet Union has tied renewed cooperation in space to a US decision to end the Strategic Defense Initiative (SDI), but there are signs that that linkage may have been dropped.

That there are scientific benefits to US-Soviet cooperation in space is undeniable, says Nancy Lubin, the author of a 1985 Office of Technology Assessment* report on the subject. But Lubin also says that, before any agreement can be renewed, the United States must more clearly define its policy objectives. How, for example, would the United States show displeasure with a Soviet action, such as the arrest of a US journalist on espionage charges, under the terms of such an agreement?

Technology transfer issues must also be clarified. Lubin says the International Traffic in Arms Regulations defines all space technology as an "implement of war". The Department of Defense is reluctant to have high-technology items leave the United States for Eastern European countries. Lubin argues that these potentially conflicting goals will have to be reconciled before an effective agreement can be renegotiated.

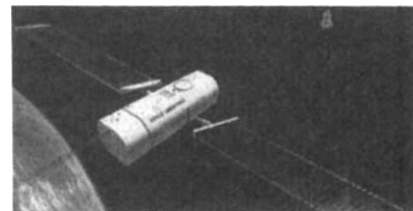
Joseph Palca

* *United States-Soviet Cooperation in Space*, Office of Technology Assessment, OTA-TM-STI-27, Washington, DC (1985).

Space facility

Washington

WESTINGHOUSE Electric Corporation has joined forces with Space Industries Inc. to produce the first privately owned manned space laboratory for research and



manufacturing. The prototype Industrial Space Facility will cost \$250 million, and will be available for launch by the space shuttle in late 1990. The facility will consist of a 2,500 cubic foot pressurized cabin attached to a 900-1,800 cubic foot supply module. The solar panels will be able to generate approximately 12 kW of sustainable power, and 60 kW for short duration needs. □

Britain in space

Soviet launch planned

THE British National Space Centre sprang to life last week with an expedition to Moscow, at the invitation of Inter-Kosmos, where a protocol was signed between British and Soviet officials to regulate the terms on which British instruments may be launched by Soviet rockets in the years ahead.

The most immediate application of the agreement is likely to be to the flight of an instrument on the Soviet X-ray satellite Roentgen, likely to be launched in the 1990s. A group at the University of Birmingham has been collaborating informally on the project for some time.

The formal protocol covers ultraviolet and X-ray satellites, solar-terrestrial physics and investigations in the infrared and microwave regions. But the British group appears to have been especially attracted by the Soviet Phobos project, designed to send a landing craft to the surface of Mars and due to be launched in 1988.

According to one member of the British group, the visit was quite separate from other visits in the past few weeks, although the members of the group were careful to emphasize that their first loyalties are to the European Space Agency. Those in the British group understand that there will be visits from scientists from other European countries in the next few weeks, especially from France and West Germany, seeking opportunities for launching scientific instruments for which launching vehicles have recently become scarce throughout the world.

John Maddox

Shroud to be dated

THE Roman Catholic Church is about to see one of its most famous relics submitted to the obvious test: pieces of the Shroud of Turin are to be taken to seven laboratories around the world for radiocarbon dating. Given the sensitivity of current techniques, less than 5 milligrams of cloth can yield a date with an accuracy of ± 60 years or so. By grouping all the laboratory results together, the statistics should be considerably better.

The Pontifical Academy of Sciences, which is responsible for the exercise, has selected five centres able to carry out dating by tandem accelerator mass spectrometry: the universities of Oxford, of Arizona in Tucson and of Rochester, New York; ETH Zurich; and the Centre pour Faibles Radioactivités at Gif-sur-Yvette, France. This technique, in which samples are pyrolyzed, ionized and accelerated to yield sensitive detections of relative isotope abundances, will be complemented by more conventional dating to be carried out at Harwell (United Kingdom) and the US Brookhaven National Laboratory.

The results of the tests will be available by Easter 1988. Typically such dating can be carried out in less than a month. A sampling room is to be built at Turin, however, and the investigators are also giving themselves enough time to handle their samples with the care appropriate for such rarities.

Philip Campbell

US semiconductors

Fears for domestic security

Washington

THE health of the US semiconductor industry is not only of economic concern. Recent studies are beginning to suggest that, if the US semiconductor industry continues to lose ground to Japan and other competitors, there could be serious national security implications.

A National Research Council report* concludes that certain critical components such as liquid crystal displays, bubble memories and many raw materials are already unavailable in the United States.

So great a dependence on foreign suppliers makes the military vulnerable should imports be cut off for military, political or economic reasons.

The report suggests several solutions to the problem, including stockpiling critical materials, setting up dedicated manufacturing facilities and redesigning equipment to make use of domestic components. But William Hittinger, chairman of one of the National Research Council committees that wrote the report, says the most pressing need is for a better descrip-

tion of the size of the problem. Now, Hittinger says, there is no database that can be used to determine accurately just how dependent is the military on parts unavailable in the United States.

The report encourages the Department of Defense to consider investing in the development of specific components for use by the military. The Pentagon is already involved in programmes to develop very high speed integrated circuits (VHSICs) and gallium arsenide technology (specifically microwave/millimetre-wave monolithic integrated circuits). Although such dedicated programmes have their place, Hittinger believes they tend to lag behind advances driven by the marketplace, making a healthy industry all the more important.

The National Security Council has also asked for a study of the Pentagon's dependence on foreign parts. An interagency group has been meeting weekly at the White House throughout the summer to determine the government's role in alleviating the problem. Darryl Garrett of the Office of Science and Technology Policy, who has been chairing the meetings, says there are some who worry that the semiconductor industry is in a precarious state. If it should go, says Garrett, "some fear telecommunications and supercomputers will be next".

Garrett himself is concerned that his group may be addressing a problem "5 or 10 years after anything can be done about it".

The Security Council report should be finished later this fall. Another report on the topic is being prepared for the Defense Science Board by a special committee chaired by Norman Augustine, president of Martin Marietta.

Ken Flamm of the Brookings Institution questions whether the semiconductor industry's economic vitality is inextricably linked to national security. Hittinger agrees that the US industry is capable of building a substitute for virtually any component temporarily unavailable in the United States, but says that, even so, the US military could be left with some difficult "work-around" problems.

The US electronics industry will probably welcome the upsurge of interest in its vitality. William Krist, vice-president of the American Electronics Association, says that one aspect of the problem he would like to see addressed is that of the Defense Department's export controls. Krist says that these have had deleterious effects on the industry's international competitiveness. Says Krist, if the Pentagon wants to do something to help domestic electronics companies, his reaction is "Hallelujah! It is about time".

Joseph Palca

Soviet Union

Corruption continues

MR Mikhail Gorbachev's campaign against graft and nepotism in the Soviet Union is not proving easy. The Academy of Sciences of Kirgizia is one of the latest objects of high-level criticism for faults ranging from nepotism and personal extravagance among top personnel to falsification of research achievements. But, according to two recent letters in the Moscow weekly *Literaturnaya Gazeta*, the presidium of the academy has not replied to the charges and, although the president of the academy, Academician Murzabek Imanaliev, has resigned, he was immediately installed as director of the academy's Institute of Mathematics in violation of the statutes requiring that such appointments are decided by secret ballot.

Imanaliev, it appears, was personally responsible for many of the abuses. One of the charges against the academy during the Kirgiz Communist Party Congress last January was that, at the previous congress five years earlier, Imanaliev had reported the production for the first time of powders of iron alloyed with rare earths, obtained by a simple enrichment process using local ore deposits. According to his 1981 report, a pilot plant to produce such powders had already been established at the Kirgiz mining and metallurgical combine. But, despite generous funding and considerable publicity, it now turns out that the specimens produced gave no evidence of the presence of rare earth elements.

Imanaliev had tried to build up the prestige of his academy by hosting all-Union symposia and conferences, often at considerable expense and with massive spending on alcohol, and outfitting offices with lavish Western furniture (disguised in the academy's account as "equipment").

Under his administration, moreover, appointments in the academy and also in the Kirgiz State University were decided not by the statutory secret ballot, but by patronage. Such appointees to academy jobs included his wife, brother and several

in-laws. Those who protested against the atmosphere of corruption got short shrift — a scholar who reported that one of Imanaliev's protégés had plagiarized his work found himself suspended from lecturing and banned from publishing in the Kirgiz SSR for five years.

According to the rules of the Communist Party, such a situation should never have been allowed to develop. The member of the secretariat of the Central Committee of the Communist Party of Kirgizia responsible for science and learning should have seen what was happening in the academy and taken appropriate steps. Instead, the incumbent of that post, Comrade Karypkulov, took advantage of the situation to advance his own academic career.

Between 1978 and 1986 he defended his doctoral thesis, published five books and appeared as author or special editor of 16 special scientific studies. He also managed to get himself elected as a corresponding member of the academy, over the heads of better qualified candidates. His removal from party office, last year, as an early victim of the Gorbachev reforms, signalled the first enquiries into academy malpractice.

Many questions still remain unanswered. In particular some unknown patronage must lie behind how Imanaliev's original appointment as president of the academy, particularly as his presidency dates from the very day he was elected to full membership of the academy (15 June 1979).

Moreover, Imanaliev's "illegal" appointment as head of the Mathematics Institute, immediately after his "voluntary" resignation from the presidency, indicates that, in spite of Party and press criticism, patronage is still a major factor in Kirgiz academic life. Imanaliev and Karypkulov were clearly supported by a major infrastructure of corruption which seems still to operate.

Vera Rich

* Foreign production of electronic components and army systems vulnerability. National Academy Press, Washington, DC 1986.

Cost of RGO move to Cambridge

SIR—There has been much discussion in your columns about the proposed move of the Royal Greenwich Observatory (RGO) from Herstmonceux to Cambridge. As I organized the main meeting in London last June, attended by many of the country's leading astronomers as well as by representatives of the Science and Engineering Research Council (SERC), I should like to comment on the present situation.

SERC intends to proceed with the move. Cambridge does not want it (as was made clear at the June meeting) and the University of Sussex does not want to lose RGO. The move is opposed on scientific grounds by virtually all astronomers, as it would mean dismantling the archives and library, putting the telescopes out of use (which would mean that new equipment could not easily be tested) and shifting the satellite laser ranger equipment. RGO would soon lose its independence. But apart from this, there is the financial aspect.

SERC has stated that my comment that the cost would be high, and would have to be borne by the taxpayer, is "misleading". In that case, why does it not produce the actual figures? The building at Cambridge would cost at least £6 million (and is not yet authorized by the university) and the sale of Herstmonceux would not raise anything like that amount. Even in the long term, the move could not be self-financing. This, at least, is the universal informed opinion, and one is bound to suggest that SERC is not releasing any figures simply because it does not know them.

If the move is authorized without a full, detailed knowledge of the cost involved, there will be severe repercussions, some of them political. SERC must therefore be compelled to make a proper statement if it intends to continue with the present plans.

It is, of course, possible that having realized the extent of its error of judgement, SERC is now casting around for a face-saving formula. One can only hope so, because the loss of Britain's Royal Observatory would be a scientific as well as a cultural and educational tragedy.

PATRICK MOORE

Farthings, West Street,
Selsey, Sussex, UK

How to verify

SIR—In his sceptical review of *Verification: How Much Is Enough?*, Dennis Fakley seems to conclude that there certainly is not enough now (*Nature* 322, 218; 1986).

Fakley claims that verification must be concerned with demonstrating an absence

of non-compliance. In other words, if it is not possible to detect with certainty every possible violation of a treaty, the treaty is, in the words of another verification sceptic, "fatally flawed". It is this type of requirement for verification that opponents of arms control hide behind.

President Kennedy defined the level of verification required as follows: the ability to detect a violation that would threaten US security in sufficient time to allow an adequate US response. This is surely a more effective definition than requiring proof of "an absence of non-compliance".

Does Fakley require the police to demonstrate that there is no crime before he feels their job is being adequately done? Arms control agreements are not religious writ to be validated absolutely. As Paul Warnke, former US arms control negotiator, said of the SALT II treaty: "We left in a little wiggle room" — in other words, disputes over minor "violations" are often differences in interpretation of a loosely worded phrase.

To claim that the Soviets are immune from international public pressure is to ignore their response to the Chernobyl disaster. Western scientists at the International Atomic Energy Agency conference in Vienna expressed satisfaction with the level of Soviet information on Chernobyl, and that is clearly due in part to the international outcry following the disaster. No country is immune from public opinion, as much as some of our leaders seem to wish it were so.

MIKE KELLY

Illinois Freeze Voters,
17 N. State Street, Suite 908,
Chicago, Illinois 60602, USA

Criteria of science

SIR—R.J. Berry, professor of genetics and member of the General Synod of the Church of England, wishes his fellow scientists to consider seriously the matter of miracles¹. He reminds us of the letter to *The Times*² in which he and fellow believers insisted: "It is not logically valid to use science as an argument against miracles. To believe that miracles cannot happen is as much an act of faith as to believe that they can happen."

If one takes as one's measure of the criteria of science those adumbrated by Karl Popper in *The Logic of Scientific Discovery*³, the point made by Berry *et al.*² cannot be gainsaid. There is no way that science can determine that tomorrow the Sun will not make abrupt movements and plunge zigzagging down onto people, then stop and zigzag back again, as was claimed by many thousands of eye-witnesses on 13 October 1917 at Fatima in Portugal. Curiously, equally devout Christians throughout the rest of the world did not notice this erratic behaviour of the Sun on that day.

In practice, of course, a scientist makes the assumption that the relationship of our planet to the Sun, which has been pretty uniform for the past 4,500 million years, is unlikely to change tomorrow — barring miracles, naturally. J.B.S. Haldane put the matter succinctly in the preface to his book *Fact and Faith*. "My practice as a scientist is atheistic. That is to say, when I set up an experiment I assume that no god, angel or devil is going to interfere with its course; and this assumption has been justified by such success as I have achieved in my professional career. I should therefore be intellectually dishonest if I were not atheist in theory, at least to the extent of disbelieving in supernatural interference in the affairs of the world. And I should be a coward if I did not state my theoretical views in public." It is useful to use the term "atheist" in the normal sense of the Greek prefix, hence "amoral", "atypical", so that "atheist" means not someone who positively asserts the non-existence of God, but rather someone who is not a theist⁴.

But this is not exactly the issue for Berry *et al.* who "gladly accept the virgin birth, the Gospel miracles and the resurrection of Christ as historical events"². Now, once it is claimed that an event has happened, it immediately comes into the purview of the historical sciences such as geology, palaeontology, astronomy and archaeology. These disciplines routinely deal with sequences of unique events and have developed their own methodologies for examining such events.

When a supposed unique event is described from the distant past, one of the first things a palaeontologist does is to try to find further examples in the record. Now when this is done in the case of the Resurrection, we find an embarrassment of riches. The number of religions in which the god-king dies and is resurrected three days later is legion. Christians even name this spring fertility festival Easter, after Oestre the Earth Goddess, and celebrate with such fertility symbols as eggs and rabbits.

Contrary to what Berry *et al.* imply, there are *scientific* disciplines for dealing with unique past events, and also for understanding the survival and role of ancient beliefs in modern society. Their willingness to accept the biblical miracles as "historical fact" says much for this faith but does not inspire confidence in their objectivity as scientists.

BEVERLY HALSTEAD

Departments of Geology and Zoology,
University of Reading,
Reading RG6 2AB, UK

1. Berry, R.J. *Nature* 322, 321–322 (1986).
2. Berry, R.J. *et al.* *The Times* 13 July 1984 p.15 (1984).
3. Popper, K.R. *The Logic of Scientific Discovery* (Hutchinson, London, 1959).
4. Haldane, J.B.S. *Fact and Faith* (Watts, London, 1934).
5. Flew, A. *The Presumption of Atheism*, (Elek/Pemberton, London, 1976).

Calculating melting temperature

Some journals still allow their contributors leisure, which makes for a readable account of a calculation of the Earth's inner core.

HAVE the journals in general, even *Nature* perhaps, become so preoccupied with what is novel that their readers cannot hope to understand what is offered to them without first rushing to a textbook, a review or to some other device for catching up? The answer is almost certainly that they have. The question whether the convention is an impediment to the spread of knowledge is not answered as easily. People who need to find out are prepared to take the trouble; usually they succeed.

Nevertheless, some journals still fly in the face of what seems to have become the general conviction that all readers at all times know everything that has happened up to now, so that each of them can hit every first paragraph running, so to speak. One of these is the *Geophysical Journal of the Royal Astronomical Society*, which from time to time lets its contributors write chapters of textbooks as yet unwritten. One of the more interesting essays of this kind is an article by J.P. Poirier of the Institut de Physique de Globe in Paris (*Geophys. J.R. astr. Soc.* **85**, 315; 1986) which combines an account of a calculation of the temperature at the inner solid core of the Earth with a critical account of the dislocation theory of melting. Poirier's article is only 13 pages long.

The need for some discussion is clear. Even if the composition of the inner core and the surrounding molten outer core were known with certainty (which they are not), an accurate calculation of the relationship between the two would be beyond the scope of experimental investigation; the pressure and the temperature are too great. So it is necessary to bring all possible evidence to bear on some assumption of what the material of the inner and outer core consists of, to use the macroscopic properties of the Earth (mass, radius, moment of inertia and so on) to estimate the density as a function of depth and then to fix conditions at the boundary between the inner and outer cores by attempting to calculate the melting curve of the material there, the relationship between its melting temperature and the pressure. But conditions near the centre of the Earth are so far away from experimental ken that the faceless equations of thermodynamics, which assume that evidently variable quantities are constant, simply do not apply. At some stage, only a model of the phenomenon of melting will help.

But what model? It is curious how the Lindemann theory of melting, essentially an extension of Debye's theory of lattice vibrations in the early 1920s, is usually the first to come to mind. (Lindemann, Churchill's chief scientific adviser during the Second World War who was afterwards made Lord Cherwell, would probably prefer to have been remembered for having learned how to correct stalling in First World War aircraft.) But this is the crudest device, based on the assumption that solids will melt when atomic vibrations become so robust that atoms are forever colliding with each other. Its thermodynamic error is that it takes no account of the free energy of the molten state.

The essence of melting is that it must be a cooperative phenomenon, one in which the whole or at least a substantial part of a melting solid is involved. For that is what the thermodynamics imply: at melting, there are two phases in equilibrium (one a solid and the other a liquid) whose energy density may be substantially different, as different as the latent heat, and which must be compensated for by the difference of entropy or of disorder between the two phases. Lindemann's theory might have been made into a more persuasive argument if, for example, he had sought to calculate how one collision between a pair of atoms in the vibrating lattice of a solid will increase the chance that one of the pair will promptly engage in another collision.

Poirier's tack is a dislocation theory of melting (there is an infinity of variations) based on the view that solids accumulate a greater density of linear dislocations in their lattice structure as they approach their melting point. That the occurrence of dislocations within an otherwise regular solid is a dynamic process is, of course, well attested; give a collection of atoms enough energy and one of them will surely lodge in the wrong place, distorting the patterns of its neighbours. The classical dislocations of the 1930s, involving the displacement of successive atoms in a line from the "true" positions, are energetically more easily formed. But the greater the density of dislocations, the less the sense of order within a lattice and the more like a liquid it appears on a microscopic scale. The extreme view is that a liquid is merely a solid saturated with dislocations.

The interest of Poirier's mini-textbook on the subject of dislocation melting is that he puts previous attempts at a comprehensive theory of melting in a sceptical

perspective. Most people seem to have recognized that, apart from the energy embodied in an isolated dislocation, there is also an energy of interaction between them, but only some models have taken account of the entropy arising from different arrangements of configurations. The physical strain within a solid occasioned by the presence of dislocation works both ways, first by increasing the volume of a solid on the way to melting but also by reducing the strain energy of later displacements. And some people, it seems, seem to have been content to leave out of the account the change of volume there must be when a solid melts.

The point of all this is to choose the most suitable dislocation theory of melting for a calculation of the equilibrium between the inner and the outer core of the Earth; Poirier settles for a theory due to Ninomiya, published in 1978. For what it is worth, Lindemann's approximation stands up well, but the theory is also nicely verified by its predictions for the properties of melting iron, certainly the predominant material in the inner solid core, under laboratory conditions.

What of the core itself? The best estimate seems to be that the temperature at the inner core boundary, that between solid nearly pure iron (with nickel as the principal impurity) and the outer molten core (which may have lighter elements as impurities) is within 100 K or so of 6,150 K. This value is in the range covered by previous estimates, and for the time being there is probably little to choose between them except that Poirier provides such a detailed account of how, by his calculations, the melting temperature should change with pressure so that it should be possible to wring some relevant information from the seismic measurements.

The picture of the inner Earth that emerges from the calculations is not very different from that now widely accepted. Both at the boundary between the mantle and the core (at a depth of just under 3,000 km) and at the boundary between the inner and the outer cores (at a depth of just over 5,000 km) there are temperature discontinuities. The former, a drop of several hundred degrees K between the liquid outer core and the base of the mantle, is easily enough accounted for by latent heat. That there should be a jump as large between the outer and the inner cores is more surprising, but consistent.

John Maddox

Cognition

Parallel distributed processing

from Stuart Sutherland

THE human mind is sloppy and prone to error, but it has the saving grace that its mistakes are usually small ones. Until recently, computer models of intelligence exhibited none of these characteristics. They operated with great precision, and when corrupted, whether by a hardware or a software fault, gave rise not to small errors but to nonsense. Recently there has been much interest in a new way of simulating the mind that reproduces some of its sloppiness: it is called parallel distributed processing (PDP) or connectionism. One example of this type of processing is described by D.E. Rumelhart, G.E. Hinton and R.J. Williams on page 533 of this issue.

The PDP model postulates a set of units with properties similar to those of neurones. They are connected together with varying strengths, and are arranged in layers. Usually, there is an input and output layer and there may be intervening layers. In such systems the use of a given concept is represented by the activation in parallel of particular sets of units within each layer. A given unit may be active when very different concepts are entertained, though of course the pattern of active units differs from one concept to another. Because many units may take part in the representation of a single concept or mental process, small errors are likely to appear in the output as new learning changes the connectivity of some of the units. Again, because a single concept is not represented by a single unit, the system may continue to function after the loss of some units; in an analogous way, human memory often survives local damage to the brain tolerably well.

Such systems also generalize in a natural way. For example, if a system has learned to give a particular output to one item it will tend to give the same output to a new item that resembles the original one. Indeed, it has been shown that the more a new item resembles a prototypical item in a class, the more likely the system is to classify it as belonging to that class, just as people classify a robin as a bird more readily than they do a penguin. A further striking result is that the systems behave as though they are acting according to rules without the rules being explicitly represented, for the system's behaviour depends merely on the pattern of distributed connections within it. It is as though new instances were treated by analogy with old ones without any explicit rule being formed. This may well be the way children learn a grammar — they certainly have no conscious rule for deter-

mining what is the subject of a sentence and they may well have no unconscious representation of such a rule. Perhaps they merely treat new sentences as analogous to others they have encountered.

Apart from the psychological plausibility of PDP, it is clearly consistent with what is known of the nervous system. Moreover, its parallel architecture means that a great deal of processing can be achieved in a very short space of time. Synaptic delays and transmission time along axons restrict the brain to taking at most 50 serial steps in a quarter of a second, but this is long enough to recognize complex objects, an act that requires a vast amount of computation.

Quite apart from their naturalness, it has been found that PDP systems can simulate some of the more surprising findings about cognition. One system, for example, designed to learn how to form the past tense of a verb, makes much the same errors in the learning process as do children. Young children start by learning the commonest verbs, which are frequently irregular — “do, did”; “come, came”. Children next master rarer and more regular verbs like “call, called”. At this stage the child often begins to make mistakes with the irregular verbs it has already learned, saying for example “he camed” or “he comed”. At the final stage, children become proficient in their use of both regular and irregular verbs. The computer model, although not designed to do so, mimics the children's performance at each stage. Several other PDP simulations have produced results remarkably similar to the behaviour of people. For example, a model simulating the learning of words came to distinguish between letter strings that conform to the pattern of English spelling but were not real words (for example, slet), and strings that could not possibly be words (for example, strz).

One of the problems in constructing such models has been that there was no adequate method of changing connectivities in a network with three or more layers. A possible solution is proposed in the report by Hinton and colleagues in this issue. The authors take the difference between the actual output and the desired output for a given input and use a rule to change the strengths of the connections to the output in the direction needed to give the desired output. This procedure is then reiterated at each level of the system. He demonstrates that the recognition of symmetry and the learning of family trees can be achieved by this technique. One difficulty with learning systems of this sort is

that optimal performance may never be reached because the system settles into a state in which the connectivity yields imperfect performance but in which any small changes to it only make performance worse (a local minimum). The learning rule of Hinton *et al.* yields good but not necessarily optimal levels of performance, but there is no proof that this would be true under all conditions.

Although for reasons not stated, they do not regard their technique as “a plausible model of learning in brains”, it could be an important step forward. Whereas learning to associate two or more things together (for example, a person's face and his voice) can be achieved using PDP by existing learning rules that operate only at two levels, such rules do not apply to learning that depends on using the results of the outputs (knowledge of results): it is likely that here some form of back propagation must take place, and this is what Hinton and his colleagues have achieved. The problem is, however, often much more complex than are the authors' simulations. In a golf swing, the output is a series of muscle movements but the knowledge of results is too often the sight of one's ball being sliced into a ploughed field. How can this visual input gain access to the motor control system that produced the slice or change the strength of the connections that determined the swing, particularly when the exact strengths used on a given swing are probably no longer available by the time the ball is seen? The problem of the mechanisms by which knowledge of results, or even reinforcement, affects performance is an important one and has been largely overlooked by psychologists, and the ideas of Hinton *et al.* may provide a lead.

Regardless of this, the naturalness and power of PDP suggest that it is here to stay. Although some members of the old guard are resistant to these ideas, the doyen of the psychology of cognition, Professor George Miller, has remarked that it is the most important revolution in psychology in his day and his day includes the advent of cybernetics, information theory, generative grammar and the digital computer as a tool for simulating the mind. His pronouncement should not be taken lightly. Although there are many unsolved problems in PDP the existing results have been achieved by only a handful of workers. As the interest and importance of these systems become more widely known we may expect the young and flexible to explore their implications in detail, using no doubt the sloppy analogies that characterize both the human mind and PDP and that are almost certainly the foundation of creative thinking. □

Stuart Sutherland is Professor of Experimental Psychology at Sussex University, Brighton BN1 9QG, UK.

Exotic atoms

Antiprotons tickle nuclei

from C.J. Batty

WITH a few exceptions, atomic physics (concerning the orbitals of particles about the nucleus) and nuclear physics (concerning nuclear excitations) rarely meet. But in the experiment of W. Kanert *et al.* (*Phys. Rev. Lett.* **56**, 2368; 1986) they have clearly met head on, with the X-ray transitions of antiprotonic atoms interacting resonantly with nuclear excitations.

Kanert and his colleagues used antiprotons from the low energy antiproton ring at CERN, near Geneva, to generate 'exotic' atoms. In such atoms, one of the electrons is replaced by a heavier, negatively charged particle, such as a muon, pion, kaon or antiproton, and is formed by stopping the particles in suitable target material. Because of the larger mass of the trapped particle, its orbits will be much closer to the central nucleus than those of the remaining electrons. Because there is only one heavy particle, the Pauli exclusion principle is not involved and the

lap with the nucleus. As the energies that would excite the nucleus itself are frequently comparable to the energies of the muonic atom, the nucleus can be excited during the muonic cascade. In the case of negative hadrons like the antiproton, which interact strongly with the nucleus, the hadron will be absorbed by the nucleus while at a relatively high n -value, where most of its wavefunction is far outside the nucleus. The paradoxical result is that in general the probability of nuclear excitation in hadronic exotic atoms is extremely small. However M. Leon (*Nucl. Phys. A* **260**, 461; 1976) pointed out that in certain special cases, where the atomic de-excitation energy is closely matched by a nuclear excitation energy, a resonance condition is satisfied and large nuclear excitation effects may be observed. When this occurs the exotic atom can be considered as being partly in the atomic state (n, l) with the nucleus

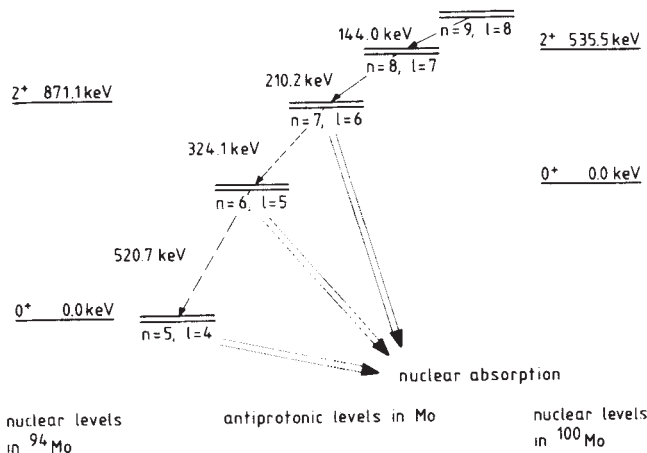


Fig. 1 Nuclear and antiprotonic atom levels in ^{94}Mo and ^{100}Mo . The populations of certain levels diminish rapidly through the absorption of the antiprotons into the nucleus. (From Kanert *et al.*)

full range of classical orbits characterized by the principal quantum number (n) and angular momentum l ($0 \leq l \leq n-1$) are available. As a result, the outer electrons can generally be ignored and the exotic atom has many properties closely similar to those of the simple, one-electron hydrogen atom. The exotic atom is initially formed in an excited state of high n and then de-excites by a mixture of Auger (electron emission) and X-ray transitions, with the latter predominating in the latter stages of the cascade. When the so-called classical 'circular' orbits, with $l = n-1$, are strongly populated so that the X-ray transitions are generally of the type $(n, l = n-1) \rightarrow (n-1, l = n-2)$.

In muonic atoms, the muon wavefunction for low n has a considerable over-

lap with the nucleus. As the energies that would excite the nucleus itself are frequently comparable to the energies of the muonic atom, the nucleus can be excited during the muonic cascade. In the case of negative hadrons like the antiproton, which interact strongly with the nucleus, the hadron will be absorbed by the nucleus while at a relatively high n -value, where most of its wavefunction is far outside the nucleus. The paradoxical result is that in general the probability of nuclear excitation in hadronic exotic atoms is extremely small. However M. Leon (*Nucl. Phys. A* **260**, 461; 1976) pointed out that in certain special cases, where the atomic de-excitation energy is closely matched by a nuclear excitation energy, a resonance condition is satisfied and large nuclear excitation effects may be observed. When this occurs the exotic atom can be considered as being partly in the atomic state (n, l) with the nucleus in its ground state and also partly in the $(n-2, l-2)$ state with the nucleus in an excited spin-parity 2^+ state. Because of the stronger overlap of the $(n-2, l-2)$ atomic wavefunction with the nucleus, absorption of the hadron from this state results in significant induced absorption from the (n, l) atomic state, so reducing the intensity of the $(n, l) \rightarrow (n-1, l-1)$ X-ray. Leon referred to these exceptional cases as 'ticklish nuclei'. Leon

predicted that the effect should be strong in antiprotonic atoms of the molybdenum isotopes and it is these which were studied by Kanert *et al.*

Figure 1 shows the nuclear and atomic levels in ^{94}Mo and ^{100}Mo which are relevant to this resonance effect. In the case of ^{94}Mo the spacing between the antiprotonic $(n, l) = (7, 6)$ and $(5, 4)$ levels (844.8 keV) is sufficiently close to the first 2^+ nuclear excitation energy (871.1 keV) to allow mixing of the $(7, 6; \text{spin-parity } I^\pi = 0^+)$ and $(5, 4; I^\pi = 2^+)$ atomic-nuclear states. This admixture reduces the $(n=7) \rightarrow (n=6)$ X-ray intensity compared with that, for example, in ^{92}Mo where the resonance does not occur. This is precisely what is seen in Fig. 2, where antiprotonic X-ray spectra of five different Mo isotopes

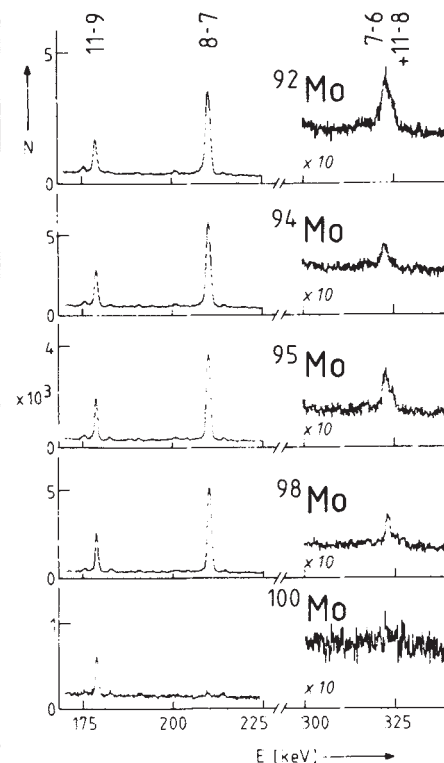


Fig. 2 Antiprotonic atom X-ray spectra for various Mo nuclides: counts per channel (N) versus energy (E). (From Kanert *et al.*)

measured by Kanert *et al.* are shown. Normalizing the intensity of the $(n=11) \rightarrow (n=10)$ X-ray transition, the $7 \rightarrow 6$ transition has a measured intensity of around 11 per cent in ^{94}Mo compared with a value of 40 per cent in ^{92}Mo .

The effect is seen even more dramatically in ^{100}Mo . Here the spacing between the $(8, 7)$ and $(6, 5)$ atomic levels (534.3 keV) is so close to the first 2^+ nuclear excitation energy (535.5 keV) that there is a very strong coupling between the $(8, 7; 0^+)$ and $(6, 5; 2^+)$ atomic-nuclear states. As a result the $8 \rightarrow 7$ X-ray transition is strongly attenuated (Fig. 2) with an intensity of around 5 per cent in ^{100}Mo compared with 122 per cent in ^{98}Mo where there is no resonance.

Why are these results of particular interest? First, the unique and predicted strong coupling in antiprotonic ^{100}Mo has been observed for the first time; indeed the coupling is so strong that the $9 \rightarrow 8$ X-ray line is also broadened and this too is observed experimentally. Second, in the case of ^{94}Mo the $(n=5, l=4)$ level is normally hidden by the very strong absorption from the $(6, 5)$ level that prevents direct X-ray transitions to the $(5, 4)$ level from being seen. This experiment provides information about the characteristics of the $(5, 4)$ level and gives an important test of models of strong interaction effects in antiprotonic atoms. $\square$

C.J. Batty is at the Rutherford Appleton Laboratory, Chilton, Didcot, Oxford OX11 0QX, UK.

Medical research

Are we losing the war on cancer?

from Marie M. Cohen and Jared M. Diamond

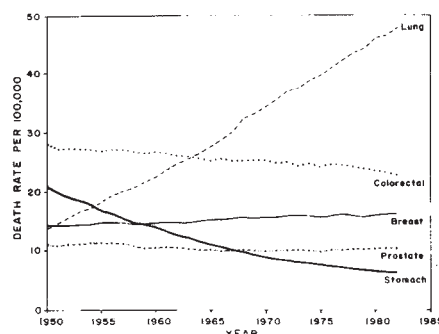
CANCER is the second leading cause of death in the United States, and to defeat it, the government in 1971 declared a 'war' to be led by the National Cancer Institute (NCI) of the National Institutes of Health (NIH). Since then, NCI has spent \$15 billion, its current annual budget is close to \$1 billion (the largest of any NIH institute), and other government and private agencies devote \$1 billion more to cancer. What progress has been made?

There have been impressive advances in curing certain cancers, notably most childhood cancers, Hodgkin's disease and certain leukaemias and testicular cancers. But these cancers are rare compared with those of the lung, colorectum, breast, prostate and stomach. The overall picture in the United States is summarized in NCI publications (see, for example, ref. 1) and in a recent controversial article by Bailar and Smith². Relative 5-year survival rates (cancer patient survival adjusted for deaths from other causes by comparison with a similar age distribution from the general population) have improved slightly from 47 to 49 per cent between 1973 and 1978, the most recent years analysed. Despite this, age-adjusted mortality rates (see figure) have been increasing because age-adjusted incidence rates have increased. Thus, by the least equivocal criterion — mortality — we are losing the war on cancer.

Some disagree with this view. "The national statistics are inevitably a few years behind the times and therefore do not reflect the most recent advances in treatment"³. "The implication that the war on cancer is being lost is ridiculous. There are many cures already there. They just haven't come out of the computer"⁴. But this latter view, although often advanced in the past decade, "has never been vindicated by national statistics when these eventually became available"³. If predictions of future trends should consider recent advances, they should also consider setbacks such as the long-continuing steep rise in incidence and mortality of lung cancer (see figure), which is likely to leave the overall mortality trend flat or

rising in the foreseeable future.

Cancer researchers disagree strongly about how to fight the war. One view is to continue present NCI policy, which puts more money into treatment (especially its biochemical/cellular aspects, which have undoubtedly spawned advances in treatment) than into prevention (see page 184 of ref. 1). Perhaps new drug development,



Age-adjusted mortality rates for five of the commonest cancers in the United States, 1950–1982. Age adjusted to the US population of 1980. (From ref. 2.)

monoclonal antibodies and studies of oncogenes, retroviruses and immunology will produce breakthroughs. It is agonizingly difficult to decide at what stage an expensive, long-term strategy should be considered a failure or whether it requires more time. One guide is to consider whether there are other, undersupported approaches with more promise. Several suggest themselves.

First, smoking is estimated to cause 30 per cent of cancers. It is the main cause of the rise in lung cancer since 1945 to its current rank as the commonest cancer and also contributes to cancers of the oesophagus, bladder, pharynx and mouth^{1,5}. Although NCI spends about \$20 million a year in a programme to control tobacco use, the tobacco industry spends \$2 billion on advertising and the US government spends \$3.5 billion to subsidize the tobacco industry. In Norway a national anti-smoking programme that involved banning cigarette advertisements resulted in fewer people taking up smoking. An end to subsidies and a ban on tobacco advertising would save the US government a sum nearly equal to the entire NIH budget while also promoting the war on cancer.

Second, lung cancer exemplifies a broader, self-evident point: it is better to prevent cancer than merely to try to treat it. As suggested by geographical variation in cancer incidence as well as dramatic shifts in incidence when a human group migrates from one country to another,

about 80 per cent of cancers are thought to have environmental causes, including 35 per cent that are influenced by diet^{1,5}. Changes in environmental factors, and not changes in treatment, are responsible for the two biggest recent components of change in US cancer mortality: the rise in lung cancer and the decline in stomach cancer (see figure). These facts, as well as elementary common sense, suggest that more money should be spent on cancer prevention than on cancer treatment — the reverse of present policy.

Third, for a prescribed drug to be efficacious, patients must take it. Do they? A recent study by Levine *et al.*⁶ monitored patient compliance by measuring blood levels of prescribed drugs and their metabolites after the patient had supposedly taken a dose. Patients claimed to have taken about 35 per cent of their doses, but in fact compliance was only 17 per cent, a value raised to 40–50 per cent by intensive educational and supportive programmes. Few oncologists are aware of these astonishingly low compliance rates, and even fewer devote much effort to patient education or to measuring compliance. When MOPP, a multi-drug chemotherapy for Hodgkin's disease, is said to produce a cure rate of 50 per cent, does this mean that the real cure rate is 50 per cent, or that the rate was 100 per cent but only 50 per cent of patients were compliant? Nobody knows, and there is surely a strong case for research on improving compliance with existing drugs rather than inventing new ones.

Finally, physicians tend to treat cancer as a biochemical/cellular process that either kills the patient or is cured, leaving no other consequences requiring professional attention. In fact, 'cancer' has become a group of mostly chronic illnesses in which the new patient must cope with repeated disruption or devastation of personal competence, morale, family relations, friendships, career and income⁷. There is little support for research in this area. There is also insufficient funding for professional personnel to help surviving cancer patients rebuild family relations, career and self-image, although physicians routinely make such referrals for patients surviving heart attacks.

Given the obvious advantage of avoiding cancer, why is more support given to treatment than to prevention? Part of the reason may be a confusion between the goals of basic and applied biology. The goal of applied biology is to solve prob-

Errata

In the article by Velia Fowler (*Nature* 322, 777; 1986) the figure was incorrectly attributed to ref. 4. In fact it was taken from ref. 6: Byers, T. J. & Branton, D. *Proc. natn. Acad. Sci. U.S.A.* 82, 6153; 1985.

In the article by R. Letolle (*Nature* 323, 19; 1986) the start of the second paragraph should read "In 1886 Crookes had just discovered yttrium..."

1. National Cancer Program. 1983–84 Director's Report and Annual Plan FY 1986–1990 (NIH, Bethesda, 1986).
2. Bailar, J.C. & Smith, E.M. *New Engl. J. Med.* 314, 1226 (1986).
3. Cairns, J. *Scient. Am.* 253(5), 51 (1985).
4. Holland, J. *The Cancer Letter* 12, 7 (1986).
5. National Cancer Program NCI Fact Book (NIH, Bethesda, 1985).
6. Levine, A. *et al. Proc. Am. Soc. Clin. Onc.* 3, 71 (1984).
7. Cohen, J. *et al. Psychosocial Aspects of Cancer* (Raven, New York, 1982).

lems by means of any available knowledge, whereas basic biology currently favours certain types of knowledge (for example, reductionist, molecular and obtained under controlled conditions at the laboratory bench). Research on oncogenes, viruses and immunology is highly rewarded; research on methods of coping, patient compliance and motivating people to stop smoking is scorned as 'soft', eligible for no Nobel prizes and little support. In an infinitely rich world, one would increase expenditure on prevention and

treatment, but an increase in the large US cancer budget cannot be expected when support for other areas of medicine and science is even stingier. Thus, some reallocation of funds from the biochemical/cellular aspects of treatment towards prevention and broader aspects, seems called for. □

Marie M. Cohen is in the Department of Medicine and Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024, USA.

Retroviral vaccines

How close is C to D?

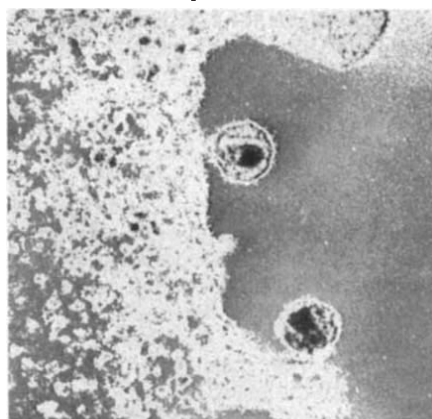
from Jerome E. Groopman

PERHAPS the highest octaves in recent meetings in Paris and Cold Spring Harbor, New York were reached in discussions of the development of retroviral vaccines. There has been considerable controversy concerning the optimal approach to the production of a safe and effective vaccine against the human immunodeficiency virus (HIV), the aetiological agent of the acquired immune deficiency syndrome (AIDS), and the significance of neutralizing antibodies. These antibodies are defined by their ability to kill the virus *in vitro* but their role in protection against infection is not clear. Marx, Gardner and their co-workers at the California Primate Research Center in Davis, in collaboration with Syntex Corporation, have now developed a vaccine based on inactivated whole virus that protects monkeys from infection and subsequent immune deficiency related to the simian AIDS serotype 1 virus (Marx, P.A. *et al. J. Virol.* in the press). Although the simian AIDS virus is a type D retrovirus whereas HIV is type C, it may help understand the host immune responses necessary for protection and the significance of neutralizing antibodies.

Lasky and co-workers at Genentech have adopted a different approach to a vaccine against HIV by cloning and expression in mammalian cells of the major viral envelope glycoprotein gp 130 (Lasky, L.A. *et al. Science* **223**, 209; 1986). Immunization of rabbits and guinea pigs with the recombinant gp130 elicits neutralizing antibodies of moderate titre that appear capable of inhibiting not only the HIV isolate used in the cloning but several other isolates as well (R. A. Weiss *et al.*, personal communication). This success in eliciting neutralizing antibodies was repeated by scientists in the National Cancer Institute and Duke University who used gp120 purified from HIV-infected lymphoid cell cultures to immunize goats (Robey, W.G. *et al. Proc. natn. Acad. Sci. U.S.A.* in the press). Similarly, F. Wong-

Staal *et al.* (personal communication) have used non-glycosylated HIV envelope fragments expressed in *Escherichia coli* and elicited neutralizing antibodies.

Although these approaches appear to follow the successful development of vaccines for viruses that do not belong to the retroviral family, the significance of the neutralization response as an indicator of



Budding of HIV at the surface of a lymphocyte (Ch. Daguet, Institut Pasteur).

success has been questioned. Furthermore, the presence of neutralizing antibodies in patients with AIDS has been taken as evidence of their inability to limit the pathogenic effects of HIV. Cell fusion and syncytia formation, perhaps caused by gp120 interaction with the CD4 determinant on uninfected lymphocytes, is thought to be a mechanism of lymphocyte loss and immune deficiency.

The approach adopted by the Davis group is simple and direct, and is modelled on the Salk vaccine for the poliovirus. Formalin-inactivated simian AIDS virus was used with the adjuvant threonyl muramyl dipeptide to vaccinate six macaques, and a control group received adjuvant alone. Moderate titres of neutralizing antibody were elicited in the vaccinated monkeys. Persistent viraemia was prevented in vaccinated animals and none developed disease, whereas five of six

controls were viraemic, with four developing clinical simian AIDS. Although the pathogenesis of simian AIDS resulting from the type D retrovirus may differ from disease related to HIV, the study from California is the first report of a vaccine which prevents spontaneous retrovirus-induced immunosuppressive disease in primates.

Recent data from our laboratory and elsewhere indicate that the humoral neutralizing response in man following HIV infection does not appear until 6–9 months after the primary antibody response. The reasons for this are not yet understood, but it is possible that the presence of neutralizing antibody before exposure to HIV might effectively limit entry and spread of the virus. The ability of human or animal neutralizing sera raised to gp120 to block formation of multinucleated giant cells *in vitro* is being studied in several laboratories. Although it is possible that the configuration of neutralizing epitopes on the viral envelope in the infected cell membrane may differ from those in native or recombinant antigens, it is still unclear whether this is of practical importance in vaccine development.

The direct approach to the development of an AIDS vaccine still has considerable hurdles, most notably strain specificity. It appears that the neutralizing antibodies raised against the recombinant gp130 and the native gp120 will neutralize some but not other geographically distant isolates of HIV (R. A. Weiss *et al.*, personal communication). Again, we can look back to the development of a vaccine against poliovirus where success was achieved after systematic serotyping of numerous strains to give the current trivalent formulation. A polyvalent subunit vaccine encompassing the major serotypes of HIV will probably be required, although how many major strains will be needed is not known.

Other approaches being considered in the development of an AIDS vaccine, such as synthetic peptides related to neutralizing epitopes of the envelope protein and anti-idiotypic antibodies, as well as study of the immune response to proteins other than gp120, should be pursued. Considering that 5 million or more individuals throughout the world are estimated to be infected with HIV, and the continued rapid spread of the retrovirus in the West and in Africa, a variety of approaches should be taken. The next step, trial of prototype HIV subunit vaccines in chimpanzees, will tell us if the direct approach works. This should teach us a lesson in the alphabet of clinical retrovirology. □

Jerome E. Groopman is in the Division of Hematology/Oncology, New England Deaconess Hospital, Harvard Medical School, Boston, Massachusetts 02215, USA.

Astronomy

Black hole origins of active nuclei challenged in Armenia

from Martin Ward and Ian Robson

THE Byurakan Observatory in Soviet Armenia has made outstanding contributions to the study of active galactic nuclei. Over the past 20 years Victor Markarian (who died late last year) and co-workers have identified more than 1,500 celestial objects with ultraviolet excess. This ensemble includes many examples from the active galactic zoo — starbursts, Seyfert galaxies and quasars — so the observatory provided a natural focus for a meeting*. The conference encompassed the study of nearly normal galaxies to the super energetic quasars — a range in energy output of more than a million — but some general themes emerged: for example, is material being pulled into or thrown out of the nucleus?

E. Khachikian reminded us that one of the first people to suggest that ejection of material plays a major role in galactic nuclei was academician Ambartsumian. The question of ejection from, and accretion onto, active nuclei has proved rather complex. It is well established that radio jets are associated with ejection, whereas the interpretation of optical evidence is less clear. Optical filamentary structures or jets have been seen and in some cases they strongly suggest the outflow of gas from the nucleus. On the other hand, the inflow of material is also indicated. The famous active galaxy NGC 1275 has a system of approximately radial filaments which are believed to result from cooling gas falling toward the gravitational centre of the galaxy. It is possible that this phenomenon may be generalized to explain the weak optical emission lines seen in the nuclei of elliptical and SO galaxies. Elaine Sadler (Kitt Peak National Observatory) showed that in cases where there are strong optical emission lines a correlation exists with the nuclear radio emission. In Seyfert nuclei the bulk motion of some of the high density ionized gas may exceed the escape velocity of the central massive object, while other material, further from the nucleus, may lose angular momentum and move inwards. This originally distant matter is eventually accreted by the nucleus to provide fuel for the ultimate energy source powering the activity.

The once controversial subject of quasar redshifts — a cosmological versus an intrinsic or local origin — is no longer an issue. As the conference demonstrated, the evidence in favour of the cosmological interpretation is now over-

whelming in the view of the vast majority of astronomers. Reviews by Howard Yee (University of Montreal) and John Hutchings (Dominion Astrophysical Observatory) showed that the argument has shifted to discussion of what sort of host galaxy a quasar resides within, spiral or elliptical, and whether interactions or other morphological disturbances are a common feature in the nearby examples. Here, imaging with the sensitive charge-coupled device detectors has recently much strengthened the association between quasars and parent galaxies as well as indicating that they may be found preferentially in groups of galaxies. The effects of the powerful radiation field of quasars on the interstellar material in the parent galaxy may be sufficient to light up the entire galaxy, resulting in observable spatially extended emission lines.

A new controversy concerning active nuclei has emerged over the past few years. Roberto Terlevich (Royal Greenwich Observatory) and Jorge Melnick (European Southern Observatory) propose to remove the need for the presence of exotic objects (black holes) at the centres of at least some active galaxies and replace them with superhot, massive stars which they claim are produced naturally as a result of simple stellar evolution. It turns out that the part of the continuum emission that ionizes the gas is similar to that predicted for the radiation processes associated with the black hole models. Thus,

the observed emission line strengths can be matched by both models. Furthermore, the high supernovae rate that is a consequence of the hot star hypothesis can reproduce the emission line characteristics of the more energetic Seyfert galaxies. Even the collimated radio structures observed might be produced from confined stellar winds and bipolar flow. Those favouring the presence of exotic objects have pointed to the highest luminosity quasars, which would be difficult to explain by thermal processes. Also, the very short-term quasi-periodic X-ray variations in several of the lower luminosity Seyferts are more likely to be understood in terms of an accretion process.

No-one would argue that ionization by hot stars takes place in so-called starburst nuclei. Nor that in some cases the regions around Seyfert nuclei are dominated by a young stellar population. The argument really concerns how general is the stellar origin of the ionizing continuum. The field is ripe for observational tests and some are already under way.

As a postscript to the conference a few of the participants visited the Soviet 6-metre optical telescope. This, the world's largest instrument, is now equipped with an image photon counting system for spectroscopy. In the past Soviet astronomers have sometimes been justifiably peeved by their western counterparts' apparent disinclination to reference or take full notice of some areas of their research. This situation is likely to change, as the technological gap in the field appears to be narrowing. □

Martin Ward is in the Astronomy Department of the University of Washington, Seattle 98195, USA; Ian Robson is in the School of Physics and Astronomy, Lancashire Polytechnic, Preston PR1 2TQ, UK.

Fast glacier flow

A soft bed is not the whole answer

from Gordon de Q. Robin

NEW evidence of a few metres of soft deforming glacial till beneath the low slopes but fast flow of Ice Stream B in West Antarctica^{1,2} led Boulton³ in a recent *News and Views* article to renew his call for a shift of emphasis in the study of ice sheets, past and present. He considers that glaciers find sliding over rock both harder and slower than motion over deformable till, and that such easy motion over till was common among former mid-latitude ice sheets. Although primarily directed at effects of future climatic change on ice sheets, a recent conference was relevant to Boulton's plea and brought together evidence and ideas on the flow of three types of glaciers: ice

streams; surging glaciers whose fast flow is intermittent; and tidewater glaciers whose fast flow leads to their discharge into the sea*.

What is the difficulty in understanding fast glacier flow? Almost certainly such flow implies 'sliding', but to many glaciologists the term covers both sliding of clean ice over hard bedrock and/or the rapid deformation of a basal layer up to a few metres thick, be it of ice, ice-moraine mixture or water-saturated moraine. Theoretical studies of glacier flow over such a deforming boundary layer are available^{4,5} but the problem is to define the

* Observational Evidence for the Activity in the Nuclei of Galaxies IAU Symposium 121, Byurakan, Armenia, 3-7 June 1986.

* The American Geophysical Union Chapman Conference, Whistler, British Columbia, 4-8 May 1986.

laws governing sliding and/or strong deformation within the boundary layer. I agree with Hutter *et al.*⁶ who wrote "as long as experts debate the proper form of sliding law, it is illusory to estimate longitudinal strain effects of basal stress". Until this problem is solved our modelling of ice sheets will be inadequate.

In these circumstances, the obvious need is to improve our knowledge of factors governing the flow of existing ice sheets and glaciers carrying large ice fluxes. Over much of Antarctica and Greenland, most surface motion appears to be caused by internal deformation rather than basal sliding^{7,8}. Only in fast-moving peripheral ice streams and trunk glaciers is sliding dominant^{9,10}.

In contrast to the soft till beneath the Antarctic Ice Stream B^{1,2}, the section of Variegated Glacier where surging started moves over water-saturated basal moraine almost impenetrable to drilling (W.D. Harrison, University of Alaska). The basal shear stresses were around 200 kPa before surging and only half that during surging (C.F. Raymond, University of Washington). But this should be related to a stress of only 20 kPa on Ice Stream B². The effective pressure Δp borne by the basal till itself, that is, overburden pressure minus basal water pressure, was put at 50 ± 40 kPa for Ice Stream B² while for Variegated Glacier it varied from 0 to 400 kPa during surging and 0 to 1,600 kPa at other times. Rapid changes of glacier motion were closely correlated with changes of borehole water levels and water discharge at the terminus¹¹.

Both Ice Stream B and Variegated Glacier gave evidence of very active erosion. During June 1983 the water-borne sediments discharged from beneath Variegated Glacier during the surge were estimated to be equivalent to 10 mm of rock over the whole glacier bed, and 0.5 m of rock during the entire surge cycle of 20 years (25 mm per year) (Humphrey, University of Washington). Assumption of steady-state sediment flow beneath Ice Stream B gave a corresponding figure² of 5 mm per year. Similar figures have been reported from large fast glaciers in Central Asia¹².

Basal moraine characteristics and driving stresses of tidewater glaciers appear close to that of Variegated Glacier than Ice Stream B. The correlation of short-term changes of ice movement with tidal heights near their termini shows that changes of Δp of the order of 30 kPa has a considerable effect on the movement of tidewater glaciers. Although Jacobshavn Glacier moved faster when tidal levels were higher as expected, the opposite was true of Columbia Glacier. M.F. Meier and collaborators (US Geological Survey, Tacoma) put this down to changes of longitudinal stress/strain rate against the terminal moraine — as indicated also by



100 years ago

PROFESSOR J. C. ADAMS, whose portrait we this day present to our readers, entered St. John's College, Cambridge, in 1839. He soon gave promise of those great mathematical powers that have brought such renown to his University. By what seems to have been an inspiration of genius, he was guided after taking his degree to concentrate his talents on the solution of an astronomical problem of excessive interest but of corresponding difficulty. The planet Uranus had shown irregularities in its motion. The orbit differed from the elliptic path which an undisturbed planet would pursue, and the deviations could not be fully accounted for by the influences of the other known planets. The only explanation of the discrepancy which astronomers could be expected to favour lay in the supposition that there was some other still more remote planet yet unknown.

It was the search for this unknown planet which attracted Professor Adams. We can imagine the delight with which a well-equipped mathematician would throw himself into the solution of such a problem. On it he was to concentrate the powers that had been cultivated during his University career. The discov-

ery of Neptune was a brilliant inauguration of the astronomical career of Adams. We find that he has worked at and written upon the theory of the motions of Biela's comet; he made important corrections to the theory of Saturn; he made an elaborate investigation of the mass of Uranus, to which he was naturally attracted



from its importance in the theory of Neptune; he has improved the methods of computing the orbits of double stars; but next to the discovery of Neptune the fame of Adams mainly rests on his researches on the moon and on the theory of the November meteors.

From *Nature* 34, 565; 14 October 1886.

increasing mean velocities of the terminal zone over the past decade as the zone has thinned because of ablation. Some adjustments to theory are needed. The case for the soft-bed glaciology of Ice Stream B received most support from field studies of deformation of soft till beneath the slowly moving Breidermerkur Glacier, Iceland (G.S. Boulton, University of East Anglia) and from the build up of soft till beneath the lower end of the small surging Trapridge Glacier, Yukon (G.K.C. Clarke, University of British Columbia).

Out of some 50 papers, only one presented quantitative data of past fast flow using a wealth of evidence on Vashon Glacier, a lobe of an ice sheet discharging into the Puget Lowlands of western Washington around 14,000 years ago (Hallett, University of Washington). Velocities near the 'equilibrium line' (given from dates of advance and retreat) were around 600 m per year, the driving stress was around 40 kPa and Δp was low with water separating till from ice in some as shown by sorted deposits.

Although some outer lobes of former ice sheets have characteristics similar to Ice Stream B, the case³ for using such a low basal stress model over wide areas of ice sheets is not obvious. It is more difficult to explain observed sea-level lowering during the last ice age with such a model. Alternatively, channelled water flow beneath ice sheets away from the margins would prevent water pressures building up to the level where basal till

would deform readily, as suggested by Shoemaker¹³.

All cases of the fast ice flow for which field evidence was presented at the meeting probably involve rapid shear of basal till associated with high basal water pressures (low Δp). The resultant basal shear stresses still vary widely, being low when the till is mainly composed of soft mud but ranging widely around more usual values when the basal till consisted of morainic boulders. An adequate explanation is still needed for how large quantities of glacial mud are produced, transported, then deposited to produce low friction elsewhere; and we still need a better understanding of sub-glacial hydrology, especially beneath large ice sheets. □

1. Blankenship, D.D., Rooney, S.T., Bentley, C.R. & Alley, R.B. *Nature* 232, 54 (1986).
2. Alley, R.B., Blankenship, D.D., Bentley, C.R. & Rooney, S.T. *Nature* 322, 57 (1986).
3. Boulton, G.S. *Nature News and Views* 322, 18 (1986).
4. Fowler, A.C. & Larson D.A. *Proc. R. Soc. A* 363, 217 (1978).
5. Hutter, K. *Theoretical Glaciology* (Reidel, Dordrecht, 1983).
6. Hutter, K., Legerer, F. & Spring, U. *J. Glaciol.* 27, 269 (1981).
7. Budd, W.F. & Smith, I.N. *LASH Publ* 131, 369 (1981).
8. Cooper, A.P.R., McIntyre, N.F. & Robin G. de Q. *Ann. Glaciol.* 3, 59 (1982).
9. Hughes, T.J. *Rev. Geophys. Space Phys.* 78, 1 (1977).
10. McIntyre, N.F. *J. Glaciol.* 108, 99 (1985).
11. Kamb, B. *et al. Science* 227, 469 (1985).
12. Shcheglova, O.P. & Chizov, P.P. *Ann. Glaciol.* 2, 103 (1981).
13. Shoemaker, E.M. *J. Glaciol.* 32, 20 (1986).

Gordon de Q. Robin is at the Scott Polar Research Institute, University of Cambridge, Lensfield Road, Cambridge CB2 1ER, UK.

Ocean Drilling Program

Coring the crust and the mantle

From the Leg 109 shipboard scientific party*

LEG 109 of the Ocean Drilling Program extended hole 648B in zero-age crust at the volcanically active axis of the Mid-Atlantic Ridge drilled by Leg 106 (see *Nature News and Views* 321, 14; 1986). Leg 109 planned to deepen the hole (see figure), which had been left open to 33 m below the sea floor, and to log the physical and chemical properties of layer 2 in hole 395A, located about 130 km west of the ridge axis (see figure) in 6–7 million-year-old crust. Drilled in 1976, hole 395A was one of the first holes to make use of a large conical re-entry cone, and is still one of the few deep ocean holes which penetrates more than 500 m of layer 2 basalt.

The area selected for Leg 109 drilling lies within a broad area of geochemically normal layer 2 basalt extending south about 140 km from the Kane Fracture Zone, which crosses the Mid-Atlantic Ridge at about 23°N (see figure). Detailed surveys using the submersible *Alvin* and the *Angus* deep-sea camera system have shown that typical layer 3 (gabbro) and layer 4 (peridotite) lithologies were exposed in fault blocks on the west wall of the median valley near its intersection with the Kane Fracture Zone. *Alvin* was surveying the median valley floor and walls south of the fracture zone at the same time that the drilling ship was in the area, providing an opportunity to respond at short notice to any promising sites discovered by the submersible.

On two successive attempts to begin drilling and coring hole 648B, the bottom-hole assembly was broken off. On the first occasion it was removed with a standard fishing tool, but the second failure occurred about 4 m above the re-entry cone, so the cone and sides of the hole could not be used to guide a tool over the end of the broken pipe. The problem was solved by welding a 1-m-diameter cone to the bottom of the fishing tool which was lowered over the end of the broken pipe, locked on and the bottom-hole assembly pulled up.

The core recovered in hole 648B provides a unique view of the internal plumbing system of the small volcanic cone on which it was drilled. The uppermost 30 m

consist of plagioclase and olivine phyric pillow lava, representing extrusive growth of the cone. Beneath this, there is a layer of fine-grained to glassy basalt less than 1 m thick which grades downwards into a coarsely vesicular aphyric basalt about 3 m thick. Below this, the basalt becomes massive and holocrystalline with some plagioclase-phyric intervals. This lithological sequence, combined with textural evidence in thin sections, indicates that the glassy zone must represent the quenched top of a lava pond within the cone, which probably served as a holding tank for lava being fed to flows on the valley floor. Rising gas trapped beneath the quenched surface must account for the vesicular unit, while the holocrystalline unit represents the more slowly cooled interior of the pond. Further laboratory studies of these samples may elucidate late-stage fractionation and extrusive processes operating at the sea floor.

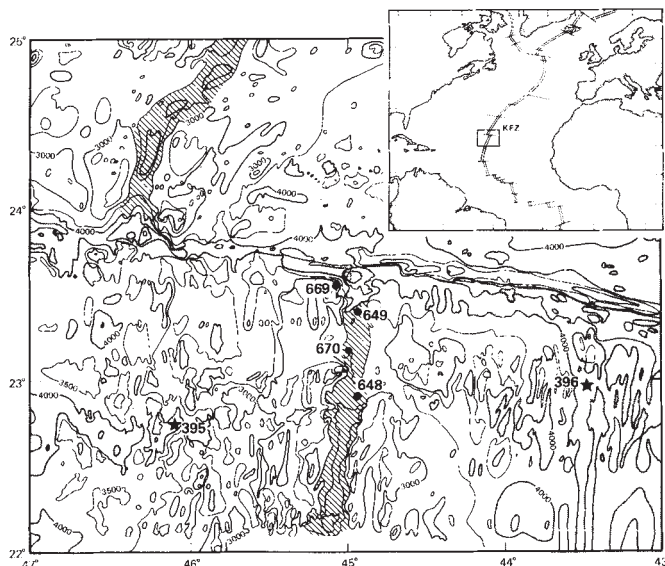
After a brief and unsuccessful attempt to spud-in a gabbro outcrop at site 669, we moved to site 395. There, we obtained an excellent set of data from hole 395A, including downhole temperature profiles; magnetic polarity and susceptibility; chemical variation; and resistivity and sonic velocity logs. One success was the packer experiment, completed previously in only one other hole, which measures permeability at different levels in the hole.

While these data from hole 395A were being collected, scientists diving in *Alvin* reported by radio that they had found a substantial area of serpentinized peridotite exposed on the lower west wall of the median valley. This seemed to be a promising place to test our spud-in capability on layer 4 material while obtaining mineralogical and textural evidence for the source and mode of emplacement of this unusual material. Site 670 was located in the centre of the area defined by the submersible survey, only 6 km from the active volcanic axis in the median valley (see figure) in about 3,600 m of water. Relying entirely

on the submersible survey, we ran pipe immediately and monitored the spud-in.

In the initial spud-in there was rapid penetration through about 7 m of sediment and rubble and into the serpentinite. When the first core was pulled, however, only half the core barrel was recovered and it was necessary to pull the pipe to recover the core and the rest of the core barrel from the coring motor. With less than four days left in the area we faced the choice of making a coneless re-entry, and continuing the drilling with the standard rotary system instead of the coring motor because the core-barrel design in the motor was obviously unreliable; or, if the re-entry failed, to spud-in a new hole with the standard rotary system.

We lowered pipe with the television monitoring the re-entry. We saw an oval



Bathymetric map of the Kane Fracture Zone showing position of drill sites. Contour intervals, 500 m. Depths greater than 4,000 m are shaded. Location of Mid-Atlantic Ridge shown by diagonal lines. Stars, sites first occupied by the Deep Sea Drilling Project; closed circles, sites first occupied by ODP. (Re-drawn from Detrick, R.S. & Purdy, G.M. *J. geophys. Res.* 85, 3759; 1980.)

pit about 2 × 3 m across filled with drilling mud, and, after a few adjustments in the positioning thrusters of the ship, the pipe moved out over the pit and disappeared into the mud. The whole re-entry took less than 15 min. We continued coring to a depth of 92.5 m below the sea floor.

The core showed that the peridotite became less serpentinized below 30 m depth. The fresher material is partially serpentinized harzburgite, containing olivine, orthopyroxene, minor clinopyroxene and spinel. Flattening and stretching of primary mineral phases, especially the orthopyroxene, produces a foliation almost perpendicular to the core. These relations suggest that the serpentinization is a consequence of its exposure at the sea floor, and that emplacement may be by low-angle transport up and out of the median valley rather than by diapiric intrusion.

*Co-chief scientists: Wilfred B. Bryan (Woods Hole Oceanographic Institution) and Thierry Juteau (Université de Bretagne Occidentale). ODP Staff scientist, Andrew C. Adamson (Texas A&M University). Also Laurie K. Autio (Univ. Massachusetts), Keir Becker (Univ. Miami), M. Mansour Bina (Lab. de Géomagnétisme du Parc St. Maur), Jean-Philippe Eissen (Centre Orstom, Noumea, New Caledonia), Toshitsugu Fujii (Univ. Tokyo), Timothy L. Grove (MIT), Yozo Hamano (Univ. Tokyo), Rejean Lebert (Univ. Laval, Quebec), Stephen C. Komor (Bureau of Mines, Avondale), Johannes Kopietz (Federal Inst. for Geosciences and Natural Resources, Hannover), Kristian Krammer (Univ. Munich), Michel Loubet (Univ. Paul Sabatier), Dan Moos (Lamont-Doherty Geological Observ.), Hugh G. Richards (Univ. Newcastle upon Tyne).

Peridotite boulders found interbedded with basalt at site 395 (Melson, W.G. *et al. Init. Rep. DSDP Leg 45*; 1978) suggests that the relationships observed at site 670 were common, at least over the past 7 million years. At present, talus moving down the wall of the median valley from

the peridotite exposure is being periodically interbedded with basalt flows from the axis of the valley only a few kilometres away. Such areas of thin or missing layer 2 (and possibly missing layer 3) seem to be common at slow-spreading centres such as the Mid-Atlantic Ridge. □

Neurology

Pharmacological treatment of central nervous system injury

from Bernhard A. Sabel and Donald G. Stein

UNTIL recently it was believed that brain and spinal cord injuries cannot be medically treated. Now, however, it seems that some of the secondary deterioration of damaged brain and spinal cord tissue can be reduced by new drugs*. These new approaches hold considerable promise for therapy and raise hopes that brain and spinal cord injury in humans can be repaired.

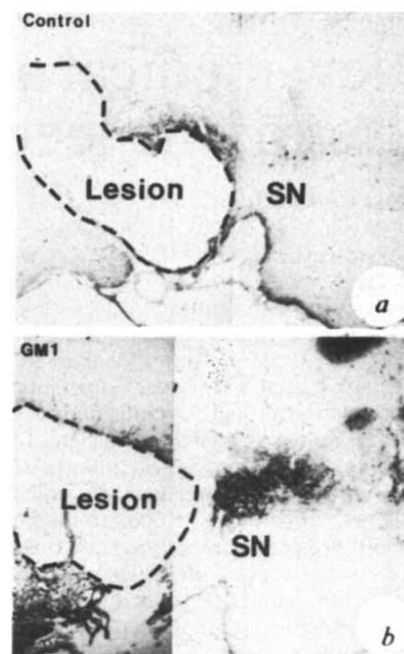
When central nervous system injury occurs, areas of the brain are lost, interrupting the circuitry of neuronal connections. Within minutes after the initial trauma, the remaining tissue starts to deteriorate further, characterized by the progressive death of partially damaged neurones near the lesion, apparently caused by a hostile biochemical environment. Current drug therapies are aimed at reducing these detrimental secondary events. The metabolites of free fatty acids contribute to the deterioration of the surviving tissue and there has been some success in reducing elevated prostaglandin and thromboxane levels in patients (Paul Demediuk, Veterans Administration Center, San Francisco).

Such pharmacological approaches are clinically successful only if the neurones spared by the initial trauma and saved from secondary cell death by drugs can maintain behavioural function. After spinal cord trauma, only 10 per cent of the axons in the ventral funiculus are sufficient to sustain near normal locomotion (M. Beattie, Ohio State University College of Medicine). It is therefore not surprising that profound early behavioural deficits are followed by considerable recovery¹. The situation is similar in spinal cord ischaemia where only a small fraction of neurones needs to be saved for the animal to be able to walk. Here, serotonin antagonists reduce both the propagation of ischaemic damage and the subsequent high mortality rate in rabbits (Justin Zivin, University of California, San Diego). By improving transmitter release of spared neurones with a catecholamine-

rgic agonist it was possible to reduce discrimination learning deficits after large neocortical lesions in rats and cats (D. Feeney, University of New Mexico).

Although all these drug treatments aim to help the surviving neurones, there are other cells of the nervous system whose role in brain repair has long been neglected: the glia cells. It has always been assumed that these cells, especially astrocytes, are detrimental to brain repair because they form a barrier around the wound, thus preventing regeneration across the injury site. But the reputation of astrocytes is improving (Manuel Nieto-Sampedro, University of California, Irvine). They have the unique ability to proliferate in response to trauma and provide an environment that can help the spared neurones to deal with the consequences of injury. For example, they buffer the neuronal environment by regulating ionic balances and binding excess excitatory amino acids. In addition, astrocytes release trophic factors that can enhance neuronal growth and survival². Because the need for neurotrophic factors is not satisfied until several days after injury, the initial balance between neurotrophic and neurotoxic substances (such as free fatty acid metabolites and excitatory amino acids) is unfavourable and cells die within several days after the primary damage. Thus, pharmacological treatments are currently being tested that try to engineer a post-injury environment with high neurotrophic and low neurotoxic levels.

Two agents that have now been tested extensively are nerve growth factor^{1,3} and gangliosides (glycolipids)⁴. Although it has been assumed that nerve growth factor and gangliosides enhance regeneration in the brain, their behavioural effects now seem to be caused by their ability to reduce secondary neuronal degeneration. It is believed that nerve growth factor is a trophic factor itself, whereas gangliosides may function as a potentiator of endogenous trophic factors. Intracerebral injections of nerve growth factor enhance neuronal survival in the septum after damage to the choli-



After transections of the nigrostriatal pathway, neurones near the lesion lose their axons and subsequently degenerate. Note the absence of cell labelling in the substantia nigra (SN) as seen here in a cross-section through the brain (a). When treated daily with GM1 ganglioside, the survival of these neurones is significantly improved. b. Dopamine-containing neurones (dark area) shown by tyrosine hydroxylase immunocytochemistry (courtesy of Drs Toffano and Consolazione).

nergic system by fimbria-fornix transections⁵. Similarly, systemic injections of purified gangliosides reduce behavioural deficits after transections of the nigrostriatal pathway⁴ (see figure). Because they act by rescuing dying neurones, perhaps via a reduction of cerebral oedema⁵, the positive effects of these drugs depend on the presence of some spared neurones⁶.

Rather than enhancing regeneration of axons, drugs help the remaining neurones to survive. Because these neurones are already committed to the function impaired by the lesion, their salvation by drug treatment appears to be the most fruitful approach to reduce the devastating behavioural deficits seen after brain or spinal cord injury. □

1. Finger, S. & Stein, D.G. *Brain Damage and Recovery* (Academic, New York, 1982).
2. Nieto-Sampedro, M. & Cotman, C.W. in *Synaptic Plasticity* (ed. Cotman, C.W.) 407 (Guilford, New York, 1985).
3. Hefti, F. *J. Neurosci.* **6**, 2155 (1986).
4. Sabel, B.S. *et al.* in *Brain Plasticity, Learning and Memory* (eds Will, B. *et al.*) 481 (Plenum, New York, 1985).
5. Karpiak, S.E. & Mahadik, S.P. *J. neurosci. Res.* **12**, 485 (1984).
6. Toffano, G. *et al. J. neurosci. Res.* **12**, 397 (1984).

Bernhard A. Sabel is at the Institute of Medical Psychology, University of Munich Medical School, Goethestr. 31, 8000 Munich, FRG and Donald G. Stein is at the Department of Psychology, Clark University, Worcester, Massachusetts 01610, USA.

*Pharmacological Approaches to the Treatment of Brain and Spinal Cord Injury Walferdange, Luxembourg 7-11 July 1986.

Invertebrate biology

New medical and scientific uses of the leech

from Charles Lent

USING leeches for bloodletting reached a medical apogee in the eighteenth and early nineteenth centuries. As everybody knows, the demise of this disgusting practice this century was a long-overdue triumph of science over superstition. Just when we thought that this slimy worm had disappeared from the physician's *vade mecum*, new therapies using leeches are emerging. Plastic surgeons use medicinal leeches to remove blood from post-operative occlusions, a procedure that increases the success of tissue transplants, reduction mammoplasty and the surgical re-attachment of amputated extremities and digits by reducing the frequency of necrosis. The chemicals secreted in leech saliva, important to these surgical procedures, are also being explored as therapeutic agents against several diseases, including atherosclerosis, thrombosis and cancer.

Leech biologists congregated recently* to discuss the medical and scientific uses of the leech, primarily the European species

*British Association of Leech Scientists Swansea, Wales, 10–12 July 1986.

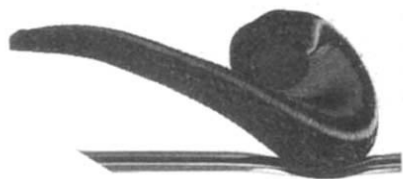


Fig. 1 A hungry leech (*Hirudo medicinalis*) attached to the side of an aquarium. The two suckers of this annelid are easily distinguished. The large posterior sucker, adhering to the glass, is used in crawling. The smaller sucker at the head encloses the mouth, which has three jaws. This anterior sucker has an array of sensory structures used in feeding behaviour (courtesy of J. S. Flannery).

Hirudo medicinalis L, and its habit of feeding on mammalian, particularly human, blood¹. Hungry leeches (Fig. 1) bite through the skin of mammals with their three jaws and make tri-radiate incisions (Fig. 2). Although both mammalian temperatures (my own work) and chemicals will evoke biting, the Na⁺ and arginine found in blood are required for leech ingestive behaviour (E. Elliott, University of North Carolina). Leeches ingest massive blood meals of 900 per cent of their body weight, and their satiation often lasts for a year. Serotonin-containing neurones, including the well-characterized Retzius cells, are obligatory for the expression of leech feeding behaviour². Skin temperatures excite synaptically these neurones to release serotonin into the periphery, where in turn it stimulates the pharynx to pump blood, the jaws to bite and gland cells to secrete saliva (Fig. 3). The salivary glands consist of large secretory cells (W. Wuttke & M. S. Berry, University College, Swansea) which are electrically independent of one another, fire Ca²⁺-dependent action potentials³ and are sti-

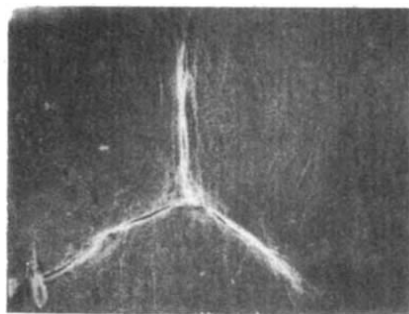


Fig. 2 A single leech bite, consisting of a set of three incisions, one from each jaw (approximately 1 mm). This bite was made in parafilm on a 37°C surface, a method for assessing leech hunger. However, leeches make similar wounds in the skin of their prey, and both kinds of bites resemble the Mercedes-Benz logo.

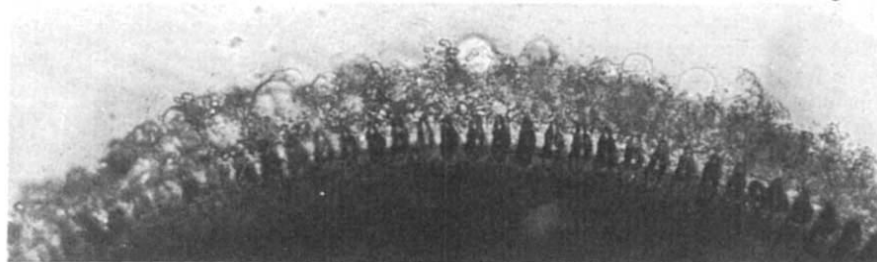


Fig. 3 Salivary secretion from a leech jaw. This jaw was isolated in leech physiological saline with many of its gland cell bodies attached. The secretory response, which is seen as a halo of saliva along the cutting edge of the jaw, was evoked by a 20-min exposure to 1 μ M serotonin. A similar response is seen when the serotonin-containing Retzius cells fire high-frequency impulses. The saliva is secreted into ducts between the paired horny cutting teeth (about 60–70 pairs in each *Hirudo* jaw) and contains mucins and the anticoagulant hirudin.

mulated by serotonin to secrete their complex biochemicals. Leech feeding in the laboratory is similar to that in nature, where gut distension rather than optimal foraging is used (J. Wrona, University of Calgary).

Hirudo are most fascinating when feeding on humans. The colourful history of leeching, which pre-dated Christ by at least two centuries (A. Young, Lydiard Millicent, Swindon), was followed by modern case histories of its use in reconstructive surgery in Europe (P. Mahaffey, Withington Hospital, Manchester) and the United States (J. Upton, Harvard Medical School). After transplanting or reattaching tissues, venous return sometimes fails and reduces arterial supply, nearly always resulting in tissue necrosis. Mechanical and chemical attempts to prevent necrosis after venous failure have been unsuccessful. The leech removes the excess blood from occluded tissues and circumvents necrosis, which provides the necessary time (about 1 week) for capillaries to grow across the sutures. Leeching has a surprisingly low risk of infection (J. Cooper, Royal College of Surgeons, London) and leeches inoculate anticoagulants that are therapeutically beneficial, causing the bites to bleed for many hours after detaching.

Chemicals in leech saliva (Fig. 3) constitute an evolutionarily derived armamentarium against mammalian haemostatic clotting mechanisms. An anti-coagulant, hirudin, was described in leech saliva a century ago⁴ and the DNA encoding this protein was cloned recently by biotechnologists at Transgène⁵. Hirudin, a 65-amino-acid peptide that functions as anti-thrombokinase, is the most powerful natural anti-coagulant known (R. Sawyer, Biopharm, Swansea). The array of active agents in the saliva of several leech species includes a hyaluronidase (J. Edwards, Biopharm), a collagenase and two fibrinases. One fibrinase disrupts clots (R. Monroe, Morrison Hospital, Swansea) and the other, perhaps, atherosclerotic plaques (R. Baskova, Moscow State University). Salivary extracts from the giant leech *Haementeria* interfere dramatically with the metastatic growth of lung tumours and inhibit tumour cell collagenase (G. Gasic, University of Pennsylvania). Two salivary aparaes prevent platelet aggregation by inhibiting their secretion of ATP (M. Rigbi, Hebrew University), but leech saliva does not contain anaesthetic. □

1. Dickinson, M.H. & Lent, C.M. *J. comp. Physiol.* **A154**, 449 (1984).
2. Lent, C.M. & Dickinson, M.H. *J. comp. Physiol.* **A154**, 457 (1984).
3. Marshall, C.G. & Lent, C.M. *J. exp. Biol.* **113**, 367 (1984).
4. Haycraft, J.B. *Proc. R. Soc. B* **36**, 478 (1884).
5. Harvey, R.P. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 1084 (1986).

Charles Lent is in the Department of Biology, Utah State University, Logan, Utah 84322-5305, USA.

Human immunodeficiency virus: an unacceptable term

SIR—In the 1 May issue of *Nature* and 9 May issue of *Science*, the International Committee on the Taxonomy of Viruses recommended that the LAV/HTLV-3 (AIDS) virus be renamed human immunodeficiency virus (HIV). Although two of the members refused to sign the recommendation, the term HIV has been increasingly used — not only in the scientific literature but also in the media. I would urge that this practice be discontinued and that a new name be selected quickly because an important argument against the term HIV has been neglected.

The views of the small group of patients suffering from the various genetically or constitutionally determined immunodeficiency syndromes (such as Bruton's type or common variable immunodeficiency) had apparently not been obtained. These patients are doing their best to cope with a disease that has nothing to do with sexual transmission or the other circumstances which are involved in virus contraction in many cases of AIDS. For many patients with an immunodeficiency that is unrelated to AIDS, the massive public discussion about AIDS has come as an extra personal difficulty. Not everybody with whom they are in contact will distinguish between genetically determined and virus-induced disease and the term HIV will add an unnecessary burden to their lives. It is the responsibility of the medical professions to avoid this problem.

PER BRANDTZAEG

Laboratory for Immunohistochemistry
and Immunopathology,
Rikshospitalet, 0027 Oslo 1, Norway

Screwworm eradication is what it seems

SIR—Under the heading 'Screwworm eradication, a grand delusion' Readshaw¹ appears to belittle the effects of the campaign to eradicate the screwworm fly (*Cochliomyia hominivorax*) from the southern United States and Mexico. He claims that most, if not all, of the changes in screwworm incidence in Texas can be explained on a model in which the number of cases in a season depends on the number of cases in the previous season and on the state-wide average temperature for the season.

A more detailed look at this hypothesis using the climatological divisions of Texas and the recorded screwworm incidence for each of the four seasons 1962–83 shows that most of the supposed explanatory power of this model is merely the result of the autocorrelation of the numbers of screwworms in one season and the next; temperature and rainfall only contribute a

Table 1 Relative amounts of variation (r^2) in logarithmically transformed screwworm case incidence explained by regressions on case incidence in the previous season (past cases), and on cases in the previous season and current mean temperature and rainfall (full regression). Dots are for Texas, 1962–83

Division	Season	Past cases (r^2)*	Full regression (r^2)*	Probability†	
				Temp.	Rain
Valley	Winter	0.42	0.39	0.39	0.76
	Spring	0.67	0.63	0.85	0.79
	Summer	0.77	0.94	0.04	0.30
	Autumn	0.79	0.80	0.10	0.21
Southern	Winter	0.47	0.52	0.10	0.98
	Spring	0.77	0.75	0.84	0.83
	Summer	0.83	0.88	0.03	0.99
	Autumn	0.86	0.87	0.93	0.11
South-central	Winter	0.32	0.51	0.06	0.58
	Spring	0.60	0.71	0.22	0.01
	Summer	0.86	0.88	0.22	0.69
	Autumn	0.84	0.83	0.31	0.97
East	Winter	no cases		—	—
	Spring	0	0	0.82	0.99
	Summer	0.88	0.88	0.84	0.24
	Autumn	0.91	0.93	0.03	0.06
Edwards plateau	Winter	0.27	0.38	0.90	0.09
	Spring	0.47	0.41	0.93	0.97
	Summer	0.77	0.84	0.02	0.98
	Autumn	0.88	0.87	0.81	0.53
Trans Pecos	Winter	0	0	0.2	0.76
	Spring	0.18	0.21	0.28	0.14
	Summer	0.84	0.94	<0.01	0.46
	Autumn	0.83	0.82	0.39	0.38
Low plains	Winter	0	0	0.19	0.94
	Spring	0	0	0.11	0.90
	Summer	0.81	0.87	<0.01	0.57
	Autumn	0.79	0.77	0.38	0.86

* Coefficients of determination adjusted for the degrees of freedom.

† Testing the null hypothesis that temperature or rainfall regression coefficients are zero.

Table 2 Mean winter and summer temperatures and total reported screwworm cases by Texas climatological division, 1962–83

Region	Winter*		Summer†	
	Temperature (°F)	Cases	Temperature (°F)	Cases
Low Plains	43.2	2	81.4	4,020
Trans Pecos	46.8	11	80.1	9,270
East Texas	47.6	0	80.9	2,868
Edwards Plateau	47.8	137	81.3	22,951
South Central	53.4	103	82.9	17,263
Southern	55.5	830	84.6	19,265
Lower Valley	59.7	194	83.9	3,340
Totals		1,277		78,977
Means	50.57		82.16	
Standard deviation	5.78		1.67	

* Mean temperatures for December, January, February; cases for January, February, March.

† Temperatures for June, July and August; cases for July, August and September.

small, and usually statistically insignificant, amount to the coefficients of determination (Table 1). For none of the divisions was the contribution of winter temperature statistically significant. Readshaw's finding of a significant effect of winter temperature is an artefact of pooling heterogeneous data, as is suggested by Table 2. Texas is climatologically and ecologically heterogeneous and is therefore not a useful biogeographical unit².

It has long been advocated by programme spokesmen that exceptional weather can have dramatic effects on screwworm incidence. The eradication programme in Florida was rushed forward to take advantage of the very cold winter of 1957–58³. The disastrous setback to the programme

in the southwestern United States in 1972 is attributed by Graham and Hourrigan⁴ and by Bushland⁵ at least partly to the exceptional weather.

However, this is all ancient history now that eradication has been announced⁶ in Mexico south to the Isthmus of Tehuantepec, thus cutting off the source of immigrant flies which used to frustrate the efforts of the southwestern US programme. There have been no cases in the United States since 1982⁵ and Pineda-Vargas⁷ has tabulated cases in Mexico from the first full year (1977) of sterile male production in Chiapas, Mexico, up to 1984 and these data are updated in Table 3. Does Readshaw claim that this steady suppression of the pest and its confinement to

Table 3 Data on monitoring myiasis cases in Mexico (from Pineda-Vargas⁷)

Year	Area searched (km ²)	No of larval samples		
		<i>C. hominivorax</i>	Rate*	Others†
1977	773,000	51,128	6.61	2,224
1978	878,990	43,097	4.90	1,828
1979	899,000	16,252	1.81	1,443
1980	1,056,650	3,983	0.38	2,186
1981	1,332,000	11,712	0.88	1,903
1982	1,332,000	16,187	1.22	2,345
1983	1,824,000	5,658	0.31	3,228
1984	1,825,660	2,821	0.16	3,442
1985	1,825,660	1,580	0.09	2,794

* Screwworm cases per km² × 100.

† Myiasis by other calliphorid fly species provides an index of sampling intensity.

the Yucatan was mainly due to the weather?

Readshaw suggests that undetected pockets of screwworms may exist. But the detectability of this highly mobile, colonizing species is particularly great due to the boom and bust patterns⁸ shown by its populations, and the damage it causes even at low densities. Undetected relic populations are thus unlikely to occur over seasonal periods.

Readshaw concludes by mentioning an outbreak in northeastern Mexico in 1985, tacitly suggesting that screwworms had not in fact been eradicated. But all 151 cases concerned were found in Tamaulipas and San Luis Potosi between 23 June and 10 August; all were discovered within 5 km of the main road despite searches by 102 livestock inspectors throughout the two states⁹; and it is suspected that the cases were due to the deliberate release of fertile pupae following the making of a large number of inspectors redundant in June. As of July 1986 no screwworms have been detected west or north of the Isthmus of Tehuantepec for a year¹⁰. We believe that this is eradication and that sterile male releases are primarily responsible.

The project is now in a position to plan with some confidence for the cost effective eradication of screwworm south to Panama¹¹. It would be deplorable if the extraordinary achievements of this project should be doubted on the basis of a fallacious argument, especially as it may discourage other attempts to control or eliminate major pest populations.

E.S. KRAFSUR

Iowa State University,
Department of Entomology
Ames, Iowa 50011, USA

H. TOWNSON

Liverpool School of Tropical Medicine,
Liverpool L3 5QA, UK

G. DAVIDSON

C.F. CURTIS

London School of Hygiene
and Tropical Medicine,
London, WC1E 7HT, UK

1. Readshaw, J.L. *Nature* 320, 407–410 (1986).2. Krafur, E.S. *Entomol. exp. appl.* 37, 297–305 (1985).3. Meadows, M.E. *Misc. Publ. Entomol. Soc. Am.* 62, 8–11 (1985).4. Graham, O.H. & Hourigan, J.L. *J. med. Entomol.* 13, 629–658 (1977).5. Bushland, R.C. *Misc. Publ. Entomol. Soc. Am.* 62, 12–15 (1985).6. Graham, O.H. *Misc. Publ. Entomol. Soc. Am.* 62 1–3 (1985).7. Pineda-Vargas, N. *Misc. Publ. Entomol. Soc. Am.* 62, 22–27 (1985).8. Krafur, E.S., Hightower, B.G. & Leira, L. *J. med. Entomol.* 16 470–481 (1979).9. *Weekly Screwworm Status Reports* (APHIS, USDA, Hyattsville, Maryland, 1985).10. Hoelscher, C.E. *Entomology Notes* 26, Texas Agricultural Extension Service (1986).11. Snow, W.J., Whitten, C.J., Salina, A., Ferrer, J. & Sudlow, W.H. *J. med. Entomol.* 22, 353–360 (1985).

UK release of genetically marked virus

SIR—While awaiting its full analysis, I should like to outline the recent experimental release by scientists from this institute of a genetically marked baculovirus in a 'cabbage-patch' ecosystem after consideration of the safety aspects and in consultation with relevant authorities. The parent virus, *Autographa californica* nuclear polyhedrosis virus (AcNPV), is a registered insecticide in the United States and has been the subject of extensive safety testing by the protocols of the US Environmental Protection Agency. Like other baculoviruses, AcNPV has many attributes that make it a desirable alternative to chemical insecticides; for example, the virus does not infect or harm vertebrates or plants or invertebrates, other than a limited number of moth pests. AcNPV has been used successfully for the agricultural control of caterpillar pests and has been shown not to have deleterious effects on beneficial or non-target insect (or other) species.

To genetically mark AcNPV, a non-coding synthetic oligonucleotide sequence (about 80 nucleotides long, with stop codons in all six reading frames and no ATG codons) was placed in an intergenic region of the viral genome at a position that deletion experiments have shown is non-regulatory and expendable. To make the marker harmless, its composition and place of insertion were such that it neither added to, nor detracted from, the coding or regulation of any viral gene product.

Prior to its release, the marked AcNPV was extensively tested in the laboratory with regard to its genetic stability and biological phenotype and in comparison with unmarked virus. The laboratory

analyses included replication studies involving more than 50 generations in insect tissue culture (sequential plaque assays) and successive passaging in caterpillars. By sequencing and other analyses no genetic instability of the virus, or marker, was detected. Titration of cloned marked and unmarked viruses in more than 70 species of UK Lepidoptera and selected Hymenoptera failed to identify differences in infectivity (susceptibility or 50 per cent lethal dose titres) or any other change in biological phenotype, and show that the cloned viruses have a highly restricted host range among UK moth species.

After evaluation of the data and in consultation with relevant agencies (UK Advisory Committee on Genetic Manipulation, Nature Conservancy Council, Ministry of Agriculture, Fisheries and Food, Department of the Environment) authority was given for a limited release of caterpillars of *Spodoptera exigua* (small mottled willow moth) preinfected with the marked virus. The release was undertaken in a field facility consisting of a netted compound which prevented dispersal of the host or its interaction with other insects or predators, thereby restricting dispersal and the opportunity for viral amplification and interaction.

The ecological object of the study was to measure precisely the ability of the virus to persist in the ecosystem under study. Analyses of persistence involved laboratory assays of virus on foliage and soil. These, plus the evidence for genetic stability, and restricted spread *in natura*, will be reported when the data have been analysed. The overall purpose of the project was to develop a risk assessment case study on the use of genetically engineered agents in the environment. This particular study concerned baculovirus insecticides, of which there is considerable practical experience and evidence concerning safety. As baculoviruses are not free-living agents, have a restricted host range and a good record of use in the environment, they have many advantages for such studies.

The long-term goal of this programme is to make custom-designed viral insecticides that, unlike broad-spectrum chemicals, are specific and limited in their host range as well as limited in their ability to persist or to have ecological impact. Since these studies represent a new era of environmental research, we believe that it is essential that a cautious developmental approach is used involving controlled field experiments that are preceded by extensive laboratory analyses. This approach will allow assessment of risk and environmental consequences to be made in a systematic manner and permit the testing of predictions concerning ecological impact.

DAVID H.L. BISHOP

Natural Environment Research
Council Institute of Virology,
Mansfield Road, Oxford OX1 3SR, UK

Learning from the apes

E.W. Menzel Jr

Gavagai! Or the Future History of the Animal Language Controversy.

By David Premack. MIT Press: 1986. Pp. 164. \$12.50, £12.50.

Silent Partners: The Legacy of the Ape Language Experiments.

By Eugene Linden. Times Books, New York: 1986. Pp. 247. \$17.95.

SOME popularized accounts of language-learning experiments sound like Jim Baker's yarn in Mark Twain's *A Tramp Abroad*:

Whatever a bluejay feels, he can put into language... Don't talk to me — I know too much about this thing... A jay can cry, a jay can laugh, a jay can feel shame, a jay can reason and plan and discuss, a jay likes gossip and scandal, a jay has a sense of humor, a jay knows when he is an ass as well as you do — maybe better. If a jay ain't human he'd better take in his sign, that's all. Now I'm going to tell you a perfectly true fact about some bluejays.

On the other hand, scholarly debates on the same topic bring to mind Diogenes Laertius:

Plato had defined Man as an animal, biped and featherless, and was applauded. Diogenes plucked a fowl and brought it into the lecture-room with the words, 'Here is Plato's man'. In consequence of which there was added to the definition, 'having broad nails'.

Nonetheless, in my opinion study of language learning in apes is, like ethologists' work on honeybees, among the most important empirical research that has been done in behavioural biology. The honeybee dance-language hypothesis was once considered controversial but it weathered that storm. Can the controversies that now surround ape studies be resolved? What are the issues?

Both of the present volumes are devoted to this question. Both have a streak of Plato in them. Neither is concerned with what apes do even without training on our part. Beyond this, however, the books are very different from one another.

Premack's essay is for academics. It centres on his own chimpanzee research but ranges into learning theory, cognitive psychology, linguistics, philosophy and biology. One of its tasks is to show that Willard Van Orman Quine

... has sold behavioral methods short: it is possible to make deeper analyses of meaning by strictly behavioral means than Quine has allowed for ... [If] the pessimism his position implies for the animal language inquiry can be lifted, it is not by mentalism but by a behaviorism stronger or more complete than his own [p.8].

Quine's basic thesis holds that any given term (for example, the fictional word

"gavagai") is ultimately inscrutable, and that therefore it is impossible even in principle to fully decipher any language, especially by behavioural techniques. I don't see what this thesis tells us about how people and animals manage to communicate as effectively as they do or how we can train them to communicate more effectively, but I'm no philosopher.

The techniques that Premack advocates are arbitrary tasks in which plastic tokens are treated as words. His logic here is that of Thorndikean animal psychology rather than ethology. First he attempts to define beforehand what the essence of any given psychological capacity or meaning amounts to; then he devises a test problem which could in principle be used with almost any species and says what behaviours any being would have to perform before he would grant it this capacity or meaning; and finally he collects some data and announces his current verdict. Among the problems that he tackles thus are stimulus meaning versus linguistic meaning, perception of causality, and whether or not animals are capable of iteration, recursion, syntactics and attributing intentions to one another.

Equal attention is devoted to whether the linguistic capabilities of humans differ from those of other species only in degree, or in kind. Here I believe that Premack misreads Darwin. Natural selection implies both continuity of descent and qualitatively different specializations of adaptations. Thus, arguments for differences in kind are not, as such, anti-darwinian, as Premack suggests. They become so, of course, if one also denies continuity and appeals to Creationism or to the so-called biological superiority of some species. Premack does not say where he stands here. I do wish, however, that he would use the terms "superior to" and "less intelligent than" less frequently. At the very least, such pronouncements about two-year-old humans versus adult chimpanzees, language-trained versus wild chimpanzees and primates versus non-primates should take into account the difference between specialized and general intelligence, or academic and practical intelligence.

Premack is critical of almost all past empirical studies. He sees no evidence against the argument that human language entails species-specific properties,

and argues further that these properties are largely innate. He maintains that if any fundamental similarities are to be found between humans and non-humans, the place to look for them is not in the linguistic domain, but in communication, cognition or perception. He questions whether human linguistic capacities can be accounted for in terms of their effect on reproductive fitness and challenges evolutionary biologists to prove otherwise. He concludes his book with an argument that linguistic considerations make sense primarily insofar as one relates them to a species' general mode of life, and especially its social organizations. On all but a few crucial points, entomologists might, of course, say much the same about ants or bees.

Future students of animal language should consider the problems that Premack discusses and the methods that he employs. To be indifferent towards the issues that Linden covers, however, would be to risk having one's laboratory shut down entirely. Linden does not explicitly predict or advocate the termination of all animal research, but some of his readers surely might.

Linden is a journalist who knows the ape language-training experiments at first hand. His book is for the general public. It does not review the scientific content of these studies in any detail. Instead, it focuses on the people who have conducted such research using American Sign Language, their personal interactions with their animals and with other investigators, and what eventually became of the animals. Anyone who doubts that the practice of science entails human foibles and passions will find food for thought. So too will anyone who sees animals as mere research tools rather than silent partners in research.

Linden's characterizations of individual scientists and the fates of their animals are often melodramatic. Chimpanzees are not simply sent from one laboratory to another, they are shipped off to Gulags and deprived even of their own names. Nevertheless, both the sociology of science and the ethical ramifications of science are illuminated. Particularly interesting to me was the ineptitude of language-trained animals when they were turned loose in Africa. According to the investigator who put aside her academic career to rehabilitate them, none of their training seemed relevant to real chimpanzee intelligence.

Linden has the least sympathy for more tough-minded researchers such as Herbert Terrace, Duane and Sue Rumbaugh and David Premack, and he is mistaken in saying that they abandoned their animals. This is not, however, to say that he condones Jim Bakerism either. Instead, he swings between the opinion that language-trained apes do precisely what one would

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

expect if they attached cognitive, emotional and communicative significance to the signs that they and their trainers make and, on the other hand, the judgement that, because of scientists' inability to agree on an airtight definition of language, the question of whether apes really do have true language is unanswerable and the whole area is therefore a quagmire. Thus, he also finds it hard to understand why Terrace and the Rumbaughes, who are no longer attempting to demonstrate language in apes, are nevertheless still conducting research of a similar sort. And in recounting a disagreement he had with Francine (Penny) Patterson about including in their co-authored book claims that gorillas tell jokes in sign language and recount what happened to them as infants, he concludes:

The story may well be true, but that Penny so enthusiastically promotes this interpretation does not enhance her credibility with the skeptical ... It is another manifestation of the bunker mentality: if people choose not to believe her, that is their problem not hers. She has Koko, and she knows how bright Koko is [p. 127].

A paradox here is that Linden himself states that without considerable first-hand

personal knowledge no one can appreciate what apes are capable of, or claim expertise on their behaviour. But where does personal knowledge leave off and the bunker mentality begin? It is easier to specify in others than in oneself.

My own bunker mentality holds that it would be just as short-sighted to write off this whole area of research on Linden's grounds as it would be to write off current biomedical research because of its failure to identify the platonic essence of life. The question is whether it has uncovered anything useful or illuminating about actual biological variations in actual living beings, and here (as such) I see little disagreement among empiricists. Given that a prime biological function of communication is to distinguish a member of one's own species or social group from a foreigner, it is not surprising that despite one's own best efforts to humanize apes, any self-respecting human, or ape, can still tell the difference. This might, however, be hindsight. To better anticipate the future, read Linden as well as Premack. □

E.W. Menzel Jr is in the Department of Psychology, State University of New York at Stony Brook, Stony Brook, New York 11794, USA.

Autobiography for the archives

F.J. Bollum

Contemporary Classics in the Life Sciences.* Vol. 1 Cell Biology. Vol. 2 The Molecules of Life. Edited by James T. Barrett. *ISI Press: 1986. Vol. 1 pp 368; Vol. 2 pp 282. \$39.95, £34 each volume.*

ALL events in the Universe leave records of their passing in the detritus that surrounds us. This is the raw material for astrophysicists, archaeologists and historians. From the garbage piles, burial grounds and other edifices they glean the habits, diet and habitat of our ancestors. From analysis of ambergris boulders we learn the travels and cuisine of the whale.

But the essence of scientific thought is more like a hot gas, and it rarely condenses into clear strata for core analysis. The Institute for Scientific Information (ISI) in Philadelphia was established to solidify all this scientific emanation into citation indexes and sources indexes that record authors and titles from a large number of scientific publications, all stored in a computer someplace. The weekly printouts of title pages in *Current Contents* provide the busy scientist with a mailing list for reprints, to be duly filed. Quarterly citation indexes provide nervous assistant professors with a gauge of tenure potential. Sitting at his computer terminal in Philadelphia, Eugene Garfield, founder of ISI, monitors the statistics of science and sets his nets for the "big fish".

Filtered from the sea of published-but-rarely-cited titles (and a few slippery Uncited Classics), are certain publications noted more frequently than others. It takes at least 100 citations, even in the more erudite disciplines, to cause a blip on the tube. The computer responds by writing a letter; "Dear Scientist: Your 1964 publication in *Silage Digest* was cited 169 times during the period 1965-1980 ... would you care to write a Citation Classic describing the personal and scientific events surrounding your important paper ...". About 50 per cent of the honoured scientists respond and in due course their time-adjusted view of autobiography appears as a Citation Classic in *Current Contents*.

Back in the terminal room in Philadelphia it has just become apparent that last

*The remaining five volumes in the *Contemporary Classics in Science* series, recently or shortly to be published, are: Clinical Medicine; Plant, Animal, and Environmental Sciences; Physical, Chemical and Earth Sciences; Engineering and Applied Science; and Social and Behavioral Sciences. All contain about 300 pages and cost \$39.95, £34.

week's *Current Contents* has as much historical value as an expired stock option. Where is all this detritus being stored? Well, it's all here in Philadelphia, on disks and backed up on tape. My God, what if a magnetic cloud from Three Mile Island passed over this place? Science history could be devastated. We had better print this stuff in hard covers so it can be buried in a time capsule. And that's what the *Contemporary Classics* series is all about. Seven volumes, three currently published, of Citation Classics wrapped up in hard covers at about \$40 per volume.

In his foreword to the first volume on cell biology, Josh Lederberg says that there is "... much to savour and more to ponder in the menu before us"; a polite way of saying that this is scientific ambergris. You can slice it and look for undigested remains of the scientific diet in the layers, or distil it and hope it will provide a base for some scientific evanescence. James Barrett, editor for these first two volumes, begins the process of stratification by organizing each volume into subtopics and providing some introductory insight into the scientific process under consideration. The reader is then free to browse through the flashbacks by

the authors. Insights laced with reminders from old love letters, rejection notices from funding agencies, projections on the future course of science, snide comments about certain unnamed colleagues, and a few choice words about the editors of *Nature* (wish I'd thought of that one!).

Many of the authors were clever enough to know that most highly cited papers are methodological or review papers (are you paying attention assistant professors?). Only the minority of us thought that somebody had actually read the conclusions of our classics, and of course we were wrong. Those of you who have not been notified that you have written a Citation Classic take heart, you may have written an Uncited Classic and thus you will not have to return your Nobel prize money.

Is this scientific history? Of course it is, and there's a lot more coming in the next five volumes. □

F.J. Bollum is a Professor of Biochemistry at the Uniformed Services University, 4301 Jones Bridge Road, Bethesda, Maryland 20014, USA. The opinions expressed are those of the author and do not reflect the views of the Uniformed Services University or the Department of Defense.

Ray and Raven

David E. Allen

John Ray: Naturalist. By Charles E. Raven. Cambridge University Press: 1986. Pp.506. Pbk £15, \$24.95.

RAYEN'S life of John Ray (1627–1705) must surely be one of the great biographies in the English language, at any rate of a leading scientific figure. Originally published in 1942, at a time of paper rationing and general preoccupation in other directions, it failed to attract the degree of attention it deserved; and though a second edition (in fact scarcely more than a reprint) was brought out in 1950, that was still well before the major burgeoning of the history of science as an academic discipline. The need for a reissue, preferably in a relatively inexpensive form, has long been glaring and it is difficult to understand why only now one at last has been produced — appropriately, as a volume in the Cambridge Science Classics series.

What makes the book outstanding is the perfect match of author and subject. Raven brought to the task a unique combination of strengths: he had a connoisseur's familiarity with and delight in Latin, the language in which Ray wrote exclusively and with high competence; he had an extensive acquaintance with the British fauna and flora, accumulated through a lifetime's fondness for natural history; he

had an intimate knowledge of Cambridge, its surrounding countryside and the workings of its University (as Master of Christ's College he was reputedly the model for Jago in C.P. Snow's novel *The Masters*); and he was as fully at ease with the seventeenth-century theological background as was only to be expected of a Regius Professor of Divinity.

Yet there is a hint in the book's dedication — "To all who like John Ray have sacrificed security & careers for conscience' sake" — that the identification went further even than that. As S.M. Walters reminds us in his brief introduction to this reissue, Raven wrote the book at a time when his openly pacifist views, proclaimed with all his customary vigour, had made him a controversial figure, not only within the University but even nationally. Matters never reached the point, as they did with Ray, of his having to sacrifice his University position and even his professional career, but the closeness of the parallels in their lives must have been only too ominously apparent.

Ray had been by no means overlooked, or even under-celebrated, before Raven's study appeared. It was left to Raven, however, to reconstruct in proper detail, with exemplary scholarship, the series of steps by which Ray, almost single-handed, so largely laid the basis of our knowledge of the vascular plants, vertebrates and butterflies of Great Britain (his explorations unfortunately never extended to Ireland) — with some side-forays into its geology and observations on the nature of

fossils. Besides exposing many misidentifications made by previous authors and correcting the over-valuation of the contribution made by Ray's young friend, patron and travelling companion, Willughby, Raven went far beyond his predecessors in capturing the adversities, not least the final years of ill health, against which Ray had to struggle and which help to bring him alive as a person.

One particularly noteworthy discovery, the product of a careful winnowing of Ray's plant records one by one, was a putative, overlooked further journey, through the north of England, in 1660. It was evidently this one which initiated the partnership with Willughby, with the historic result of extending Ray's interests to zoology, giving him the means to study plants and animals at first hand over a far wider area than his native East Anglia, and broadening his ambition, and his ultimate achievement, from the compiling of a national Flora to the devising of a pioneer "natural" classificatory system. Never reluctant to venture a judgement or a sharp dissection of the character of one of the many lesser performers on the contemporary scientific scene, Raven set a standard for works of this kind that was immeasurably superior to anything before.

Like most biographers, his admiration for his subject occasionally led him into claiming too much. It is highly questionable, for example, that the development of natural history as a strand in British life "is due to John Ray far more than to any other man". Similarly it is hard to fall in with his conviction that "without Ray's preliminary work there could have been no Linnaeus". Moreover, subsequent research has turned up new information which alters the background colouring in certain significant respects (notably in restoring to prominence the overlooked — because unpublished — Joseph Dandridge, described in these pages merely as "famous for his collection of birds' eggs"). But these are very minor defects, certainly not enough to detract from the overall impression of a masterpiece. It is a book which all historians of science should read and which, thankfully, they now no longer have an excuse not to own. □

David E. Allen, Lesney Cottage, Middle Road, Winchester, Hampshire SO22 5EJ, UK, is a past President of the Society for the History of Natural History and author of The Naturalist in Britain: A Social History (Allen Lane, 1976).

New in paperback

• *The Ecological Implications of Body Size*, by Robert Henry Peters. Publisher is Cambridge University Press, price is £12.50, \$16.95. For review see *Nature* 308, 474 (1984).

• *The Transforming Principle: Discovering that Genes Are Made of DNA*, by Maclyn McCarty. Publisher is Norton, price is \$5.95, £4.95. For review see *Nature* 317, 209 (1985).

A place in the Sun

G. F. J. Garlick

Amorphous Silicon Solar Cells. By K. Takahashi and M. Konagai. Translated by F. R. D. Appa. *North Oxford Academic**, 1986. Pp.225. £40, \$66.

THE ELECTRONICS revolution of the past four decades began with the invention of the transistor, and has rested squarely on the ability to produce large, flawless single crystals of semiconductors, notably silicon. The photovoltaic cell, which can convert some of the sunlight striking it into energy, is a product of the same age. Also made from large single crystal silicon, it was developed for use in space as a power source for satellites, and there is no doubt that it works. Solar cells have reached efficiencies of 15 per cent under space solar flux, and silicon cells with an efficiency of 18 per cent or more have been developed in the laboratory. However, although silicon is an abundant chemical element, silicon solar cells are expensive for general terrestrial use. This led to an extensive search for solar cells based on other semiconducting materials, including microcrystalline structures and, in the 1970s, amorphous silicon. When Carlson produced an amorphous silicon cell with an efficiency of 5.5 per cent at the RCA Laboratories in Princeton (reported in 1977), this instigated a spread of interest throughout the world.

Amorphous Silicon Solar Cells, translated from Japanese, gives an insight into progress to date, especially in Japan. Half of the chapters discuss solar power, solar cells of the single crystal variety, problems of production and cost reduction, and power production systems. This treatment gives a good background and perspective for the last three chapters which concentrate on amorphous silicon cells; but it also results in a lack of space for in-depth information on cell fabrication, and on research and development leading to higher efficiencies and ease of production.

At first sight, an amorphous semiconductor with its plethora of localized states at the electronic band edges, low carrier mobilities and many trapping states would seem to be a poor candidate for an efficient solar cell. However, as the authors show in an historical review, atomic hydrogen produced in glow discharge deposition takes up incomplete silicon bonds which reduces the intergrain conductivity and makes a p-i-n structure possible. Other techniques are also effective, such as boron inclusion. The latter represents the best cell production to date, especially

when a heterojunction to a silicon carbide "window" is included in the front of the cell, which gives more efficient transmission of light into the cell and inhibits carrier loss at the front. This technique, and the glow discharge method, came from Carlson and his co-workers.

There are features of the amorphous silicon cell that limit its performance under solar flux conditions, especially in space. The dominant one, which is not yet understood, is a photoelectronic degradation at these flux levels. However, such problems are not insurmountable. Another amorphous solid, selenium, had an instability problem — it reverts to a crystalline form with time. But effective research removed the problem and selenium went on to make a world impact in the form of a photocopying device (Xerox). The intensive efforts of the Japanese, ably described in this text, will no doubt provide a solution for the amorphous silicon problem — or the reason for

no solution — within the next few years. What the authors of this book do bring out clearly is that large area cells can be produced at cost levels well below those for single crystal cells. The achievement of efficiencies of 10 per cent or more for windowed cells at RCA puts such cells into direct competition with single crystal cells for terrestrial applications. Indeed, we already enjoy battery-less calculators using these cells.

This book, with its elegant colour plates of various cell types, power installations and other appliances, makes a very readable introduction to the amorphous silicon cell. The lack of detailed discussion is offset by references at the end of each chapter which give readers access to other sources of literature on the fascinating development of solid state devices. □

G. F. J. Garlick, 267 South Beloit Avenue, Los Angeles, California 90049, USA, is a consultant on the design of new-generation silicon and gallium arsenide solar cells for use in space.

Gauging problems

Paul Langacker

Group Structure of Gauge Theories. By L. O'RaiFeartaigh. *Cambridge University Press*: 1986. Pp. 172. £20, \$34.50.

THERE have been remarkable advances in elementary particle physics in the past 15 years. We now have a highly successful and mathematically consistent theory, known as the standard model, which incorporates or predicts all known facts concerning the strong, weak and electromagnetic interactions. Despite its successes, however, the standard model cannot be considered seriously as a candidate for the ultimate theory of Nature, mainly because it contains a large number of arbitrary assumptions and free parameters. There is therefore a great deal of current activity in developing still more fundamental theories, such as grand unified theories, supergravity, Kaluza-Klein theories and superstrings. One difficulty for students and experienced researchers alike is that the standard model and, even more so, these generalizations involve the use of increasingly sophisticated mathematics, especially group theory and topology.

This excellent little book provides one big chunk of the necessary background, namely group theory and its application to gauge field theories. The treatment lies somewhere between that of a review article and that of a textbook, and can best be characterized as a rapid tour through a broad class of topics in group theory, particle physics and gauge field theory.

The first half of the book deals with group theory. All of the standard topics are concisely covered, including elemen-

tary definitions, Lie groups and algebras, and representation theory. There is also a brief introduction to the application of global symmetries to particle physics and field theory.

The second half covers the algebraic and group theoretical aspects of gauge theories and spontaneous symmetry breaking. Several chapters are devoted to conventional topics such as the structure of gauge theories, the Higgs mechanism, anomalies, the standard model and grand unification. The most unusual parts of the book are the two final chapters, which employ all of the machinery developed earlier. These are respectively devoted to orbit theory, which is the study of the symmetry-breaking patterns induced by specific Higgs representations, and to the minimization of the Higgs potential. These topics represent difficult and underdeveloped areas of research that have previously not been adequately summarized.

This book is an excellent synopsis of the application of group theory to particle physics. It is crammed with useful information, some of which is hard to find elsewhere. There is a very good reference section and a glossary, and there are exercises for each chapter. The book is generally well written, though it is not without its share of typographical errors and an occasional undefined term, and it should prove very useful to advanced theory students and researchers both as a textbook and a reference work. Less experienced readers should be warned, however, that this is a very concise and dense account. It is not intended as an introduction to field theory and in practice also requires some previous knowledge of group theory. □

Paul Langacker is a Professor in the Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19174, USA.

*Now a division of Kogan Page, 120 Pentonville Road, London N1 9JN, UK.

Preliminary conclusions of the Royal Society and Academia Sinica 1985 geotraverse of Tibet

Chang Chengfa*, Chen Nansheng†, M. P. Coward, Deng Wanming*, J. F. Dewey§, A. Gansser||, N. B. W. Harris¶, Jin Chengwei*, W. S. F. Kidd #, M. R. Leeder**, Li Huan*, Lin Jinlu*, Liu Chengjie*, Mei Houjun†, P. Molnar††, Pan Yun†, Pan Yusheng*, J. A. Pearce‡‡, R. M. Shackleton¶, A. B. Smith§§, Sun Yiyin*, M. Ward|||, D. R. Watts¶¶, Xu Juntao##, Xu Ronghua*, Yin Jixiang*, Zhang Yuquan†

* Institute of Geology, Academia Sinica, PO Box 634, Beijing, China

† Institute of Geochemistry, Academia Sinica, Guiyang, China

‡ Department of Geology, Imperial College, Prince Consort Road, London SW7 2BP, UK

§ Department of Geological Sciences, University of Durham, South Road, Durham DH1 3LE, UK

|| Geologisches Institut, ETH-Zentrum, CH-8092 Zurich, Switzerland

¶ Department of Earth Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

†† Department of Geological Sciences, State University of New York at Albany, Albany, New York, 12222, USA

** Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK

††† Department of Earth and Planetary Sciences, Cambridge, Massachusetts, 02139, USA

‡‡ Department of Geology, University of Newcastle, Newcastle-upon-Tyne NE1 7RU, UK

§§ Department of Palaeontology, British Museum (Natural History), Cromwell Road, London SW7 5BD, UK

||| St Andrew's Hospital, Devons Road, Bow, London E3 3NT, UK

¶¶ Department of Geology, University of Glasgow, Glasgow G12 8QQ, UK

Institute of Geology and Palaeontology, Academia Sinica, Chi-ming-ssu, Nanjing, China

The 1985 Chinese/British expedition to the Tibetan Plateau attempted to solve the question of the origin of the very thick crustal rocks in this region. Continuing northwards movement of the Indian plate over the past 38 Myr has given rise to severe folding and thrust faulting, causing crustal thickening by internal deformation. Previous collisions of microplate terranes derived from Gondwanaland occurred during Mesozoic times but the Kun Lun terrane of northern Tibet was already part of Laurasia by the Carboniferous.

THE Tibetan Plateau along the traverse route is underlain by at least three continental fragments successively accreted¹⁻⁶ to the southern margin of Asia. On fossil and sedimentary evidence the Lhasa Terrane came from Gondwanaland; the Kunlun Terrane was part of Laurasia in the Carboniferous. The deformation and derivation of Cenozoic deposits suggest that thickening of the Tibetan Plateau crust can be explained by internal shortening and do not support the idea of long-distance underthrusting of the Tibetan Plateau by Indian continental crust. Neotectonic features along the northern half of the route consist of active east-west trending strike-slip faults, thrusts and folds, while along the southern half the previously reported N-S trending normal faults are most prominent.

We present here the preliminary fieldwork results of a geological traverse across the high Tibetan plateau during June and July 1985 which followed the 1,150-km-long Lhasa-Golmud highway. The purpose was to investigate the origin and chronology of successive collisions of continental fragments with Laurasian or Gondwanian provenance, and to determine how the crust of Tibet was thickened. Final interpretation of our work must await the results of palaeomagnetic, geochemical, geochronological, palaeontological, sedimentological and structural laboratory work. High-quality topographic maps at 1:100,000 scale were used, supplemented by LANDSAT and Metric camera satellite imagery. The pre-Cenozoic geology of the region is described from north to south; we divide it here into the Kunlun, Qiangtang and Lhasa Terranes, separated by the Jinsha and Banggong Sutures (Fig. 1).

The Kunlun Terrane

We treat the area north of the Jinsha Suture as a single tectonic unit, the Kunlun Terrane. Based on the presence of an ophiolite belt of late Palaeozoic or Triassic age in the A'nyêmaqên Shan 300 km to the east-southeast, the Kunlun Terrane was previously thought to be made up of two units, separated by the South Kunlun Suture². However, in the geotraverse area, there seems to be lithological and structural continuity across this zone. The oldest fossiliferous rocks seen by us in the central Kunlun, are a thick sequence consisting of turbidites, slates with horizons

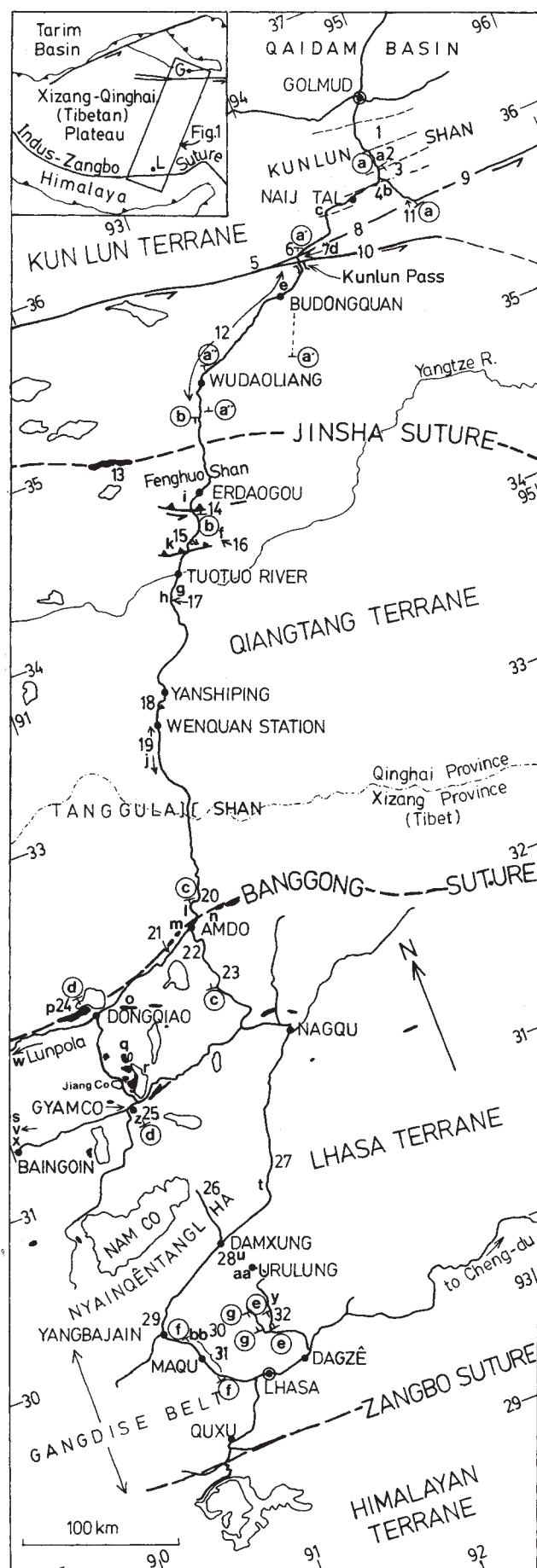
of graded limestone conglomerates, limestones and minor silicic volcanics (Fig. 1, locality 3). The limestones yield both a Lower(?) and an Upper Ordovician fauna. Further north in the Kunlun are sequences, >1 km thick, of calc-alkaline basaltic, andesitic and rhyolitic lavas (Fig. 2, column *a*), associated with fluvialite redbeds. These are overlain by Lower Carboniferous (Upper Dinantian) carbonate-clastic cycles of marine deltaic origin, which yield Laurasian coral and brachiopod faunas (Fig. 2, *a*).

Apparently unconformable on the Lower Palaeozoic sequence, although faulted against it in the central Kunlun, is a thick, coarse conglomerate containing granite and other boulders. The granites resembles those of the North Kunlun batholith (Fig. 1, locality 1). The conglomerate passes up into a mainly clastic (Fig. 2, column *b*) and volcanoclastic Permo-Triassic sequence, which locally includes 2 km or more of basaltic lavas and tuffs (Fig. 1, locality 4 and Fig. 2 column *c*) and probably originated in a rift. These are succeeded by a thick carbonate succession which ranges up to late-middle Triassic in age.

To the south, between the Kunlun and the Jinsha Suture, there is a wide zone with finer-grained, deeper-water Triassic sediments (Bayan Kala Group), consisting mainly of mudstones with graded wackes (Fig. 1, locality 12 and Fig. 2 column *e*). These are thought to represent an accretionary prism made up of turbidites and contourites, but it is possible that part of the sequence may have been deposited on a passive margin continental rise.

In the northern Kunlun, a prominent feature is an east-trending calcalkaline batholith (Fig. 1, locality 1). Early intrusions are foliated and of a more mafic, dioritic to tonalitic, composition. Later intrusions are undeformed and cut by garnetiferous pegmatites which suggests some crustal melting. This batholith intrudes, but may be cogenetic with, local Upper Palaeozoic volcanics. North of the Xidatan Fault (Fig. 1, locality 8), there is an east-trending belt of calc-alkaline orthogneisses (Fig. 1, locality 6). Around Naij Tal there are several large post-tectonic high-level granites.

Structures in the Kunlun Terrane can be separated into an early Mesozoic and a Palaeogene or younger set. We found no



evidence of major late Palaeozoic compressional deformation. The attitude and scale of the folds (Fig. 3a); the nature and intensity of the cleavage; and the magnitude of finite strains in the Lower Palaeozoic and Permo-Triassic rocks appear similar. Pebbles of Upper Ordovician limestone in the Permo-Triassic basal boulder conglomerate show no evidence of predepositional strain. Although all contacts seen are tectonically modified, the thick, Permo-Triassic basal conglomerates and the local absence of Devonian and Carboniferous rocks indicate a substantial unconformity. We believe this unconformity is caused by extensional tectonics. Therefore, the southern limit of strong Late Palaeozoic compressional deformation lies to the north of this transect.

North of the Middle Kunlun Faults (see Fig. 3a), strata are affected by open to closed folds and lack an axial-surface cleavage. Between the Middle Kunlun Faults and the Xidatan (locality 8), the very thick sequence of Permo-Triassic sediments, lavas and volcanoclastics is tightly folded, cleaved and shows internal thrusting. The structures are generally upright, but are occasionally overturned to the north. An east-west zone of post-tectonic metamorphism, locally reaching garnet grade, is superimposed on these structures. In the southern Kunlun, the most conspicuous feature is a north-dipping phyllonite zone (locality 7), probably a major detachment (Fig. 3a), which contains south-vergent shear sense indicators. South of the Kunlun Pass, steeply dipping to vertical grey-green turbidites, which young to the north, show zones of intense deformation and thrusts almost parallel with bedding. They seem to be part of a heavily imbricated thrust package. Near Budongquan, tight upright folds are seen. South of Wudaoliang, southwards facing originally recumbent refolded folds with an axial-surface cleavage are developed in Triassic flysch (Fig. 3a). They vary in orientation around upright folds with a steep axial-surface cleavage. These, in turn, are folded by more open, upright structures. These complexly deformed, well-foliated Triassic rocks are overlain unconformably by folded, but not foliated Palaeogene redbeds. A folded but not foliated sequence of possible Jurassic, locally red, clastics with coal seams, occurs in an intermontane basin (Fig. 1, locality 11) in the south Kunlun (Fig. 3a). The basal breccia of this sequence rests unconformably on marble of probable Permo-Triassic age.

The time of deformation in the Kunlun Terrane is not tightly constrained by direct evidence. However, the prominent north-derived molasse of mid- to late Jurassic age in the Qiangtang Terrane to the south (see below) brackets the event more closely. The main deformation affecting the Permo-Triassic section in the southern Kunlun seems to have begun in the late Triassic or early Jurassic.

The Jinsha Suture

This is represented in the traverse area by an inlier of ultramafics and gabbros (Fig. 1, locality 13), supposedly ophiolitic^{2,7}, which is surrounded by Palaeogene redbeds. The outcrop, located 60 km west of the road, could not be reached. However, other ophiolites, associated with radiolarites, have been recognized along strike far to the east-southeast and to the west^{2,7}.

Part or all of the Triassic flysch sequence south of the Kunlun Pass, can be regarded as occupying the suture zone. The overall arrangement of the structures in the Kunlun Terrane (Fig. 3a), the southward facing recumbency in the south and, in particular, the north-derived Jurassic molasse south of the Jinsha Suture indicate that the suture dips northwards.

Fig. 1 (left) Map of the traverse routes. Numbers along the routes locate features referred to in the text. Lower case letters, sites of stratigraphic sections shown in Fig. 2. Lines of cross-sections in Fig. 3 are shown by a T at each end with letters in circles adjacent. Black, areas of ophiolite suite outcrop (from ref. 7). Location of map shown in the inset (L and G in this inset are the positions of Lhasa and Golmud).

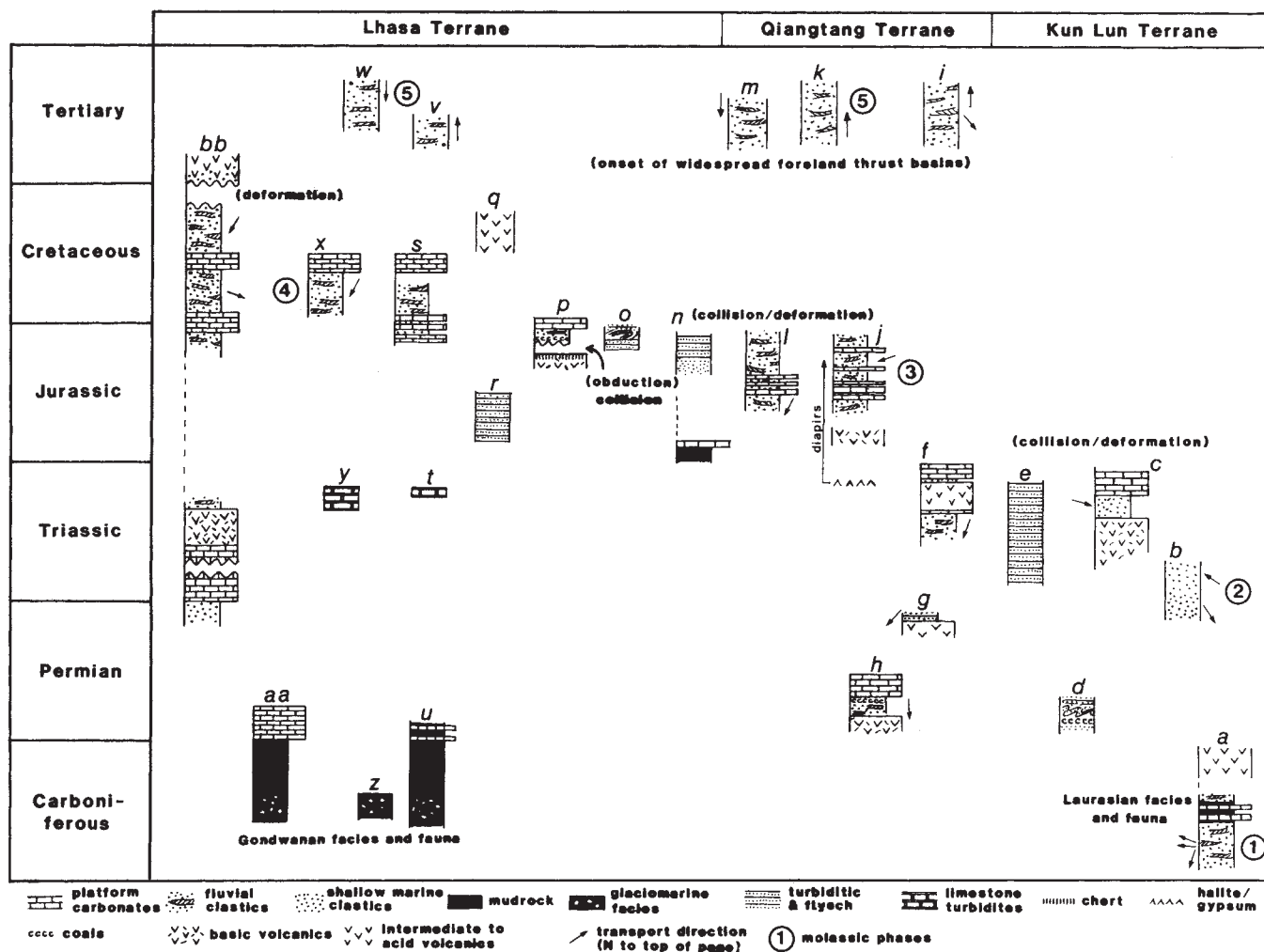


Fig. 2 Schematic sections (no vertical scale of thickness) of the more important stratigraphic sequences examined on the geotraverse. Letters (a-z, aa, bb) indicate locations, shown on Fig. 1. Ordovician sections discovered in the central Kunlun are omitted for clarity. Numbers in circles refer to phases of continental molasse sedimentation.

The Qiangtang Terrane

This terrane extends over 300 km, from the Jinsha Suture to the Banggong Suture. The oldest rocks are Permian. In the north, these include fluvio-deltaic clastics with coals and fusulinid-bearing limestones (Fig. 1, localities 15 & 17; Fig. 2 columns g & h). The flora includes *Gigantopteris*, which is typical of the Cathaysian flora. Associated basalt flows are of transitional to alkalic composition, with silicic centres marked by agglomerates and dacite-rhyolite flows. Both subaerial and submarine volcanics are present and probably indicate a rift setting. Thick Triassic fluvial gravels, derived from the north, are overlain by plagioclase-pyroxene-olivine phryic flows thought to be part of the basalt-andesite-rhyolite Batang Group which outcrops to the east of the geotraverse (16, f). Their mineralogy is consistent with a volcanic arc origin. Preliminary palaeomagnetic results from these lavas show a Northern Hemisphere palaeolatitude close to that expected for the southern margin of Asia in the Triassic. The central and southern parts of the Terrane are dominated by a thick (>2.5 km) mid to late Jurassic fluvial redbed sequence particularly well exposed near Wenquan Station (Fig. 1, locality 19; Fig. 2, column j), which interdigitates to the south with an increasingly marine facies of shallow subtidal clastics, calcareous muds and limestones deposited in an offshore shelf basin. Strongly altered mafic lavas occur locally within the Jurassic succession northwest of Wenquan Station. This great northeast derived clastic wedge (molassic

phase 3 on Fig. 2) clearly records the unroofing of the newly uplifted Kunlun Terrane.

Jurassic and older rocks of this Terrane are affected by mostly upright southeast- to east-trending folds. In the north, isolated areas of late Palaeozoic to Mesozoic rocks show upright, locally north-verging, open to tight folds lacking, or with, weak cleavage. Near Yanshiping, Jurassic redbeds are locally imbricated and folded above low angle thrusts, which probably decouple on evaporites. A large gypsum diapir has risen in an anticline ~10 km southwest of Yanshiping (Fig. 1, locality 18) and smaller examples were seen elsewhere. South of the Tanggula Shan, the folds are tighter, with a better developed axial-surface cleavage. But nowhere along the traverse in the Qiangtang Terrane are there any intensely-deformed and metamorphosed Mesozoic rocks.

The folding may be of early Cretaceous age. Locally, flat-lying red clastics of possible late Cretaceous or Tertiary age unconformably overlie folds north of Amdo (Fig. 1, locality 20; Fig. 2, column m). Near the Tuotuo River (Fig. 1, locality 15; Fig. 2, column k) folded Palaeogene red beds unconformably overlie Permian and Triassic rocks.

The Banggong Suture

Isolated outcrops of thrust-bounded ophiolites are distributed over ~180 km in the northern half of the Lhasa Terrane^{7,8,9}. The northern limit of the array is a well defined east-west line which is thought to represent the Banggong Suture between the

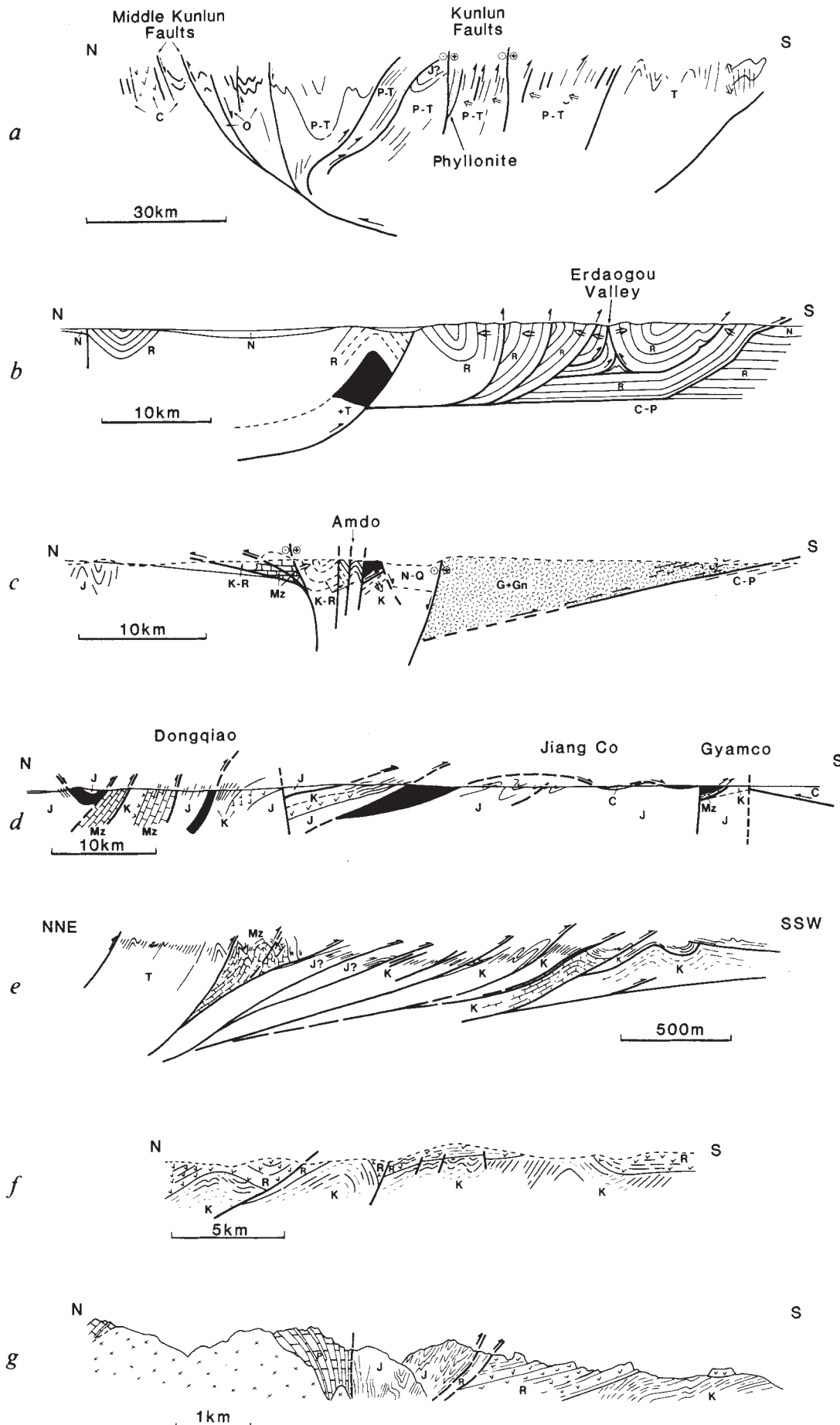


Fig. 3 Selected cross-sections of portions of the geotraverse routes. Locations shown on Fig. 1. O, Ordovician; C, Carboniferous; P, Permian; T, Triassic; J, Jurassic; K, Cretaceous; Mz, Mesozoic; R, Palaeogene; N, Neogene; Q, Quaternary; G, granitic rocks; Gn, gneisses; x, granite; v, volcanics; black, ophiolites; double-shafted arrows, younging directions.

Qiangtang and Lhasa Terranes (Fig. 1). South of the suture, between Dongqiao to Gyangco (~70 km apart), the ophiolite outcrops are closely spaced and probably represent part of a single nappe which is now disrupted (Fig. 3d). South of Gyangco, the rarer outcrops of ophiolite are gently dipping and are part of the same ophiolite nappe. Thus, the array of ophiolites represents a single flat sheet, obducted at least 180 km southwards from the suture.

The ophiolites are dismembered⁸ but are made up of a harzburgite mantle sequence, a dunite-wehrlite-gabbro-trondhjemite plutonic complex, a sheeted dyke swarm (well exposed south of Amdo (Fig. 1, locality 22)) and basalt pillow lavas. The sequence of crystallization and preliminary geochemical data support a supra-subduction zone origin. The pillow lavas are covered by Upper Jurassic deep-sea sediments.

East of Gyangco, ophiolites, originally obducted in the late Jurassic, early in the collision between the Qiangtang and Lhasa Terranes, are locally thrust over red clastics, of probable Cretaceous age. The timing of collision and ophiolite obduction is marked by a Middle and Upper Jurassic flysch sequence consisting of cleaved black mudstones with limestone turbidites which pass upwards into a thick sequence of distal clastic turbidites, and southwards into interbedded limestones and brackish clastics. This sequence (Fig. 2, phase 4) represents flexural subsidence caused by ophiolite obduction and thrust loading in the north. A few kilometres west of Dongqiao (Fig. 1, locality p) (24, p), harzburgite is unconformably overlain by soil horizons formed in a seasonally humid climate. These underlie coastal plain clastics which are overlain by cross-bedded chromite-rich fluvial sandstones. *Cladocoropsis* patch reefs of late Upper Jurassic age cap this sequence. West of Dongqiao, a calc-alkaline granodiorite/diorite pluton gives a preliminary isotopic age of ~130 Myr.

The Lhasa Terrane

This terrane extends over ~350 km, from the Banggong Suture southwards to the Indus-Zangbo Suture, which is the northern limit of exposure of the Indian continental crust. The age of basement in the Lhasa Terrane is indicated by a zircon age of 530 Myr from high-grade anatectic gneisses south of Amdo (Fig. 1, locality 23) and inherited zircon ages of ~2,000 Myr from the Yangbajain granite¹⁰. The gneisses near Amdo show a complex deformation history. An initially intense tectonic fabric was isoclinally folded in amphibolite facies, the isoclinal folds were reoriented by later east-west folds which were carried south on major biotite-grade shear zones and later brittle thrusts (Fig. 3c). Only the later stages of fold and thrust deformation are recognized in the adjacent Mesozoic rocks. No stratigraphic contacts between the gneisses and younger sediments were seen.

The oldest fossiliferous sediments in the Lhasa Terrane (Fig. 2) are Upper Carboniferous and contain a low-diversity Gondwanan fauna which includes brachiopods, bryozoans and solitary corals. Other sediments include a mixtite, containing pebbles of various granites, feldspar-phryic felsite, quartzite and other exotics in a matrix of silty mudstone which is finely laminated in places. Striated pebbles were found in two localities. Pebbles penetrating siltstone laminae must be dropstones. This mixtite, previously known¹¹ near Urulung about 70 km north of Lhasa, has now also been recognized east of Damxung (Fig. 1, locality 28; Fig. 2, column a) and near the northern margin of the Lhasa Terrane, ~10 km south-southeast of Gyangco (Fig. 1 locality 25, Fig. 2 column Z). We believe it represents a glaciomarine shelf-basin deposit (contrary to some previous opinions¹²). This interpretation and the nature of the fauna lead us to believe that the Lhasa Terrane was derived from Gondwanaland. The late Carboniferous glaciomarine deposits are overlain by shallow sub-littoral sandstones and by Permian shelf carbonates. The Permian fauna is dominated by solitary corals and contains none of the typical Tethyan fauna.

This suggests that this Terrane was still part of Gondwanaland in the early Permian.

Thick Triassic platform carbonates contain emersion surfaces, slight internal unconformities and Fe/Mn encrustations. Elsewhere, platform breakup is indicated by limestone turbidites. In the southern part of the terrane, northwest of Lhasa, Triassic(?) mainly basaltic lavas (Fig. 1, locality 30; Fig. 2, column bb) a few hundred metres thick, provide evidence for a submarine rift, which we relate to detachment of the Lhasa Terrane from Gondwanaland. Late Triassic and Jurassic reef limestones and clastics¹³ overlie the volcanics (Fig. 2, column bb) but no intact stratigraphic sections were observed passing down into the Triassic(?) lavas. A very thick (>2 km) sequence of early Cretaceous fluvio-deltaic clastics were later deposited from the north. Abundant granitoid clasts and debris are seen locally towards the north of the Terrane, around Baingoin. A widespread late Aptian-Albian transgression deposited carbonates with orbitolinids^{11,13}, and local rudist reefs. In the southern part of the Lhasa Terrane, upper Cretaceous fluvial redbeds, derived from the north, are intercalated upwards with volcanics^{12,13}, which become dominant in the unconformably overlying Palaeogene Linzizong Formation¹⁴.

In the north of the Lhasa Terrane, near Amdo, Cretaceous porphyritic granites (with lower intercept zircon ages of 140–120 Myr¹⁰) are associated with andesites and rhyolites of Aptian-Albian age (Fig. 1, locality 22). Both plutonic and volcanic rocks are probably post-collisional, and associated with the late Jurassic Banggong Suture. Alternatively, they may be related to northwards subduction under the southern edge of the Lhasa Terrane. Further south is a belt of two-mica anatectic granites (Fig. 1, locality 27) which has given a monazite age of 120 Myr from one pluton at Baingoin¹⁰.

The southern margin of the Lhasa Terrane is dominated by the east-trending 40–90 Myr calc-alkaline Gangdise batholith^{12,14} and its associated 1.5 km-thick volcanic sequence (Fig. 1, locality 31; Fig. 2, column bb) of late Cretaceous and Palaeogene andesite-rhyolite lavas and pyroclastic rocks, interbedded with continental sediments^{15,16}. The northern edge of the Gangdise Belt lies just north of Yangbajain, where a sequence of potassic lavas (Fig. 1, locality 29) was identified. Most of the Gangdise Belt traversed is attributed to magmatism at an active continental margin above a northward-dipping subduction zone^{1,2,12}. The Nyainqêntanglha Range exposes a biotite orthogneiss in contact with staurolite, garnet and andalusite bearing metasediments. The zircon age¹⁰ of <50 Myr from the orthogneiss suggests that it may represent uplifted deeper levels of the Gangdise batholith.

Most of the Mesozoic sediments and volcanics north of the Zangbo suture have upright folds, occasionally overturned and south-vergent. The steep cleavage is cut locally by a later gently dipping crenulation cleavage, but we found no evidence of the earlier structures reported by Burg *et al.*¹⁷. Mesozoic strata to the south of the Zangbo suture are intensely deformed with complexly refolded recumbent folds, unlike equivalent age strata of the Lhasa Terrane¹⁷. Representative sections for this terrane are shown in Fig. 3d–g.

At three localities in the terrane, fluvial clastics containing limestone were observed to overlie older deformed rocks with angular unconformity. Near Dongqiao (Fig. 1, locality 24; Fig. 2, column p) pre-Cretaceous obduction of the Banggong Suture ophiolite is demonstrated by this relationship⁸. In the Nyainqêntanglha range north of Damxung Fig. 1, locality 2, Cretaceous redbeds with thick basal breccias lie unconformably on strongly foliated calc-phylrites of possible Jurassic age. In both these examples, the clastics overlying the unconformity are themselves well folded and affected by thrusts, but are not cleaved to any significant extent. At Maqu (Fig. 1, locality 31) farther south near Lhasa, Upper Cretaceous red clastics with volcanics are folded, but not strongly cleaved, and are unconformably overlain by Palaeogene volcanics which are not significantly folded

(Fig. 3f). No demonstrable structural break has been found between the little foliated Upper Cretaceous redbeds seen at Maqu and the generally strongly-cleaved older Mesozoic rocks near Lhasa. If this contrast reflects different structural levels of the same deformation, the pre-60 Myr, pre-Indian collision deformation¹⁸ of the late Cretaceous at Maqu could be either a younger stage (by ~75 Myr) of collisional deformation still migrating south from the Banggong Suture, or be caused by compressional tectonics in the back-arc region of the Gangdise-Andean Arc. Alternatively, if the contrast results from an (undetected) pre-late Cretaceous unconformity (as seen in the northern Lhasa Terrane), the folding of the late Cretaceous at Maqu is probably a back-arc event for the Gangdise Arc. This late Cretaceous-early Palaeogene folding event in the southern Lhasa Terrane, if separate from the Banggong Suture-related deformation, might also include folds and thrusts that affect the mostly Cretaceous-Palaeogene redbeds in the northern area. However, some or all of these structures in the northern Lhasa Terrane may be post-Indian collision in age. In much of the Lhasa Terrane, it is not yet possible to separate structures formed before and after the beginning of the Indian-Asian collision.

Palaeogene crustal shortening

Intermediate and silicic volcanics are dominant in the Palaeogene Linzizong Formation in the southern Lhasa Terrane. Farther north, thick Tertiary fluvial redbeds fill thrust bounded foreland-type or intermontane basins. Near Baingoin, orbitolinid-rich clasts in the redbeds are derived from a north-facing thrust-scarp of Mesozoic rocks. In the nearby Lunpola basin¹⁹ (Fig. 1) northern-derived Eocene-Oligocene redbeds are strongly folded and are thrust, from the north, over more gently folded Pliocene lake beds.

Near Amdo, thrust and folded late Cretaceous(?) and Tertiary northern-derived redbeds rest unconformably on more tightly folded Jurassic rocks (Fig. 1, locality 20; Fig. 2, columns *l*, *m*, *n*) of the southernmost Qiangtang Terrane and on rocks of the Banggong Suture zone. In the northern Qiangtang Terrane, and overlapping the Jinsha Suture zone onto the southern edge of the Kunlun Terrane, strongly folded Palaeogene redbeds form a belt north of the Tuotuo River (Fig. 1, locality 15; Fig. 2, column *k*) (15, *k*) and another around and north of Erdaogou (the Fenghuo Shan) (Fig. 1, locality 14; Fig. 2, column *i*). Much of the red arenite and conglomerate forming these thick (>2 km) fluvial sequences in the northern Qiangtang Terrane is derived from the south, but a proportion has northerly provenance. The existence of these foreland-type and ramp valley basins, as well as the deformation of the sediments is evidence for lithospheric loading and shortening.

At Maqu (Fig. 1, locality 31) the Palaeogene volcanics are gently warped (Fig. 3f), suggesting¹¹ that the post-Eocene deformation is negligible. However, in this general area (Fig. 1, locality 32), prominent southwards thrusting^{11,17} of Jurassic rocks over the Palaeogene section was confirmed (Fig. 3f, g). Other post-Palaeogene thrusts, some with a component of strike-slip displacement, occur across the entire plateau, notably near Baingoin, bordering the Lunpola basin west of Amdo (locality 21), and locations 10 km and 40 km south of Erdaogou (Fig. 1, locality 14). The Palaeogene redbeds have been strongly folded across the whole of their exposure. In the Fenghuo Shan (Fig. 1, locality 14), the shortening from folding alone (Fig. 3b) is estimated to be ~40% over 30 km. A comparable shortening is seen in the Palaeogene redbeds exposed both 30 km further north near Wudaoliang, and 40 km south near the Tuotuo River. If the deformation in these fairly extensive, but isolated, redbeds is representative of the whole, the Tibetan Plateau has been shortened by at least 40% in post-Eocene times. This order of magnitude allows all the crustal thickening of the Tibetan Plateau to be explained by distributed shortening^{20,21} within the Tibetan crust. If there were no post-Eocene shortening between the Palaeogene outcrops, the estimate of overall shortening

caused by folding would be ~12%; insufficient to explain all the thickening. Because this estimate ignores large-scale thrust faulting, 12% shortening is considered to be an extreme lower limit. The low metamorphic grade of the rocks exposed along the traverse suggests that there would have to be many small displacement thrusts if thrusts contribute significantly to crustal shortening and thickening. Thrusts which are known to have moved after the start of the Indian collision (Fig. 3b, c, e, f, g) are consistent with this idea.

If large-scale underthrusting of Asia by Indian crust is proposed to explain crustal thickening²²⁻²⁶, a systematically migrating elevation front generating southern-derived molassic sediments might be expected. We found no evidence for such systematic migration and derivation of the Palaeogene and Neogene sediments. Previous palaeomagnetic studies²⁷⁻²⁹ and our preliminary results imply large amounts of crustal convergence within the southern part of Asia since the Indian collision began. The question of how this is divided between north-south shortening and tectonic escape on strike-slip faults is still to be resolved.

Neotectonics

Concordant summit levels, with rare peaks standing somewhat above that level, are a conspicuous feature of the landscape of the Tibetan Plateau. This morphology may be the product of erosion surfaces formed by scarp retreat (pedeplains). In the south, near Nagqu, such a surface seems to be locally overlain by Pliocene³⁰ deposits. In the north, near Wudaoliang, Pliocene³¹ beds also overlie eroded Palaeocene and older rocks, but the Pliocene strata here have also been involved in folding with the development of new local erosion surfaces. In most places, the ages or relative extents of these surfaces are not known, and it cannot be assumed that the surfaces are of the same or similar ages everywhere. The existence of these surfaces does not preclude significant north-south shortening by folding and/or faulting of the substrate in the Neogene—note the folded (and perhaps actively folding) Pliocene strata near Wudaoliang, and in the Lunpola Basin—nor do they imply any hiatus in convergence between India and Asia since the collision began.

The Tibetan Plateau is a region where active seismicity and faulting has influenced much of the present topography. In central and southern Tibet most of the recent structures are north-trending. Normal faults and strike-slip faults mostly trend either WNW or ENE. As demonstrated by previous work³²⁻³⁴, these structural trends are consistent with a presently dominant east-west crustal extension in Southern Tibet and the northern Himalaya. We saw no evidence of north-south normal faults north of Wenquan Station, located half-way across the Plateau (Fig. 1). North of that point, a large east-west strike-slip fault zone (described below) is the most prominent structure, but evidence for north-south compression by thrusting and folding of late Cenozoic sediments was also seen. North and south of Wudaoliang (Fig. 1), Neogene strata are warped or gently folded over an extensive area and, near Erdaogou (Fig. 1, locality 14), seem to be overthrust by Palaeogene redbeds. Probable antecedent drainage across this thrust suggests that it was active in the Neogene, and possibly the Quaternary. Farther south, in the region of the active normal faulting, young thrusting is suggested by strongly antecedent drainage across the redbed range at Amdo (Fig. 1, locality 21). Folding of Pliocene lake beds in the Lunpola basin also requires at least local north-south shortening in this southern area in recent geological time³⁵.

Two strands of the strike-slip fault approach each other near the Kunlun Pass at the southern edge of the Kunlun. Evidence for Recent slip on the Kunlun Pass Fault (Fig. 1, locality 10) is visible from near the pass to at least 25 km to the east. Small stream valleys from mountains to the north are offset left-laterally ≥ 50 –135 m. Large streams are offset more than smaller ones. An active glacier is deflected 70 m and its terminal moraine and eastern lateral moraine more than 100 m. If the valleys were

formed in the Holocene, as their morphology suggests, the slip rate is probably 5–10 mm yr⁻¹.

The Xidatan Fault (Fig. 1, locality 8) is a more important structure and continues for several hundred kilometres to the east, where it is clearly active^{36,37}. In both the Xidatan, and the Dongdatan (Fig. 1, locality 9) the fault zone is marked by prominent pressure ridges, large tension gashes (up to 2 m deep) and left-lateral offsets of small streams and fans by up to tens of metres. The large scale and the steep sides of the tension gashes imply that a large earthquake occurred on this fault within the past few hundred years. This inference is supported by the length of the zone of disruption which was observed in two segments each ~35 km in length located 35 km apart. The average displacement for this earthquake based on the offset of drainage features was as much as 10 m.

Evidence for substantial left-lateral slip on the Xidatan Fault during the Quaternary is clear. South of the Xidatan and north of the Kunlun Pass (Fig. 1, locality 7), the mountains are covered by a moraine containing boulders of Triassic slates and phyllites, granite, pyroxenite and hornblende gabbro. The moraine is overlain by imbricated stream gravels which have a northeasterly source. These gravels are overlain by varved lake sediments, northern-derived gravels and sands and an upper moraine which lacks pyroxenite and gabbro clasts. This sequence dips southwards at 10–15°. The moraines apparently came from the north³⁸, probably from mountains now located further west, north of the fault. West of the Kunlun pass and north of the fault, granite and Triassic slates and phyllites outcrop extensively. At one locality only, 30 km west of the eastern edge of the moraine, exposures of hornblende gabbro and pyroxenite were found (Fig. 1, locality 5). We infer that the moraine was left by a glacier that flowed southwards across this pyroxenite and was subsequently displaced.

The age of the lower moraine is probably Late Pliocene, or Pleistocene. From magnetostratigraphy of the overlying lake beds, it has been inferred³⁹ that the age of the moraine is ~2.8 Myr. This implies an average Quaternary slip rate of at least 11 mm yr⁻¹ for the Xidatan Fault. If both fault strands moved throughout the Holocene, a total rate of 20 mm yr⁻¹ for both in this interval seems a likely lower limit.

Conclusions

The pre-Tertiary geology of the Lhasa-Golmud traverse route defines three micro-continental fragments, the Kunlun, Qiang-

tang and Lhasa terranes, separated by the Jinsha and Banggong sutures. Faunal assemblages and a glaciomarine mixite indicate a Gondwanian origin for the Lhasa Terrane. Rifting and separation from Gondwana probably took place in the Triassic. Permian flora in the Qiangtang Terrane has Cathaysian affinities. Rifting and separation from Gondwana probably occurred in the pre-Permian. Lower Carboniferous fauna in the Kunlun is Laurasian. Pre-late Palaeozoic metamorphic basement was seen only in the Lhasa Terrane but is predicted (from the largely shelf-type sediment sequences) to underlie both the other terranes. The micro-continents were successively accreted to the southern Eurasian continental margin^{1–6}. Major deformation and detrital sedimentary wedges indicate the beginning of the collision of the Qiangtang and Kunlun Terranes in the late Triassic–early Jurassic, and of the Qiangtang and Lhasa Terranes in the middle–late Jurassic.

No evidence of late Palaeozoic compressional deformation was found in the Kunlun Terrane, which indicates, together with the fauna, that it was already part of Eurasia by the Carboniferous. Thick sequences of Palaeogene fluvial redbeds deposited in thrust-bounded intermontane basins are found at intervals across the Plateau. Strong folding and thrusting of these sequences suggests that much, if not all, of the thickening of the 70-km-thick Tibetan Plateau crust can be explained by shortening. The provenance, deformation, and distribution of these redbeds provide no support for models calling for extensive underthrusting of the Plateau by the Indian crust.

North-south normal faults and associated strike-slip faults allowing east-west crustal extension are restricted to the southern half of the traverse route. The northern half of the route reveals evidence for late Cenozoic north-south shortening and eastward tectonic escape along left-lateral east-west trending strike-slip faults. We found a Quaternary left-lateral displacement of 30 km on the Xidatan Fault and estimate a Holocene total slip rate of 10–20 mm yr⁻¹ across both strands of the Kunlun strike-slip fault.

This expedition was funded jointly by the Royal Society and Academia Sinica. Participation of P.M. and W.S.F.K. was funded by NSF grant EAR-8417640 and NASA grant NASA-G5-524. ETH Zurich contributed to the expenses of A.G. We thank Mr Len Mole of the Royal Society and Ma Xuezheng of the Chinese Academy of Sciences Foreign Affairs Bureau for their assistance.

Received 3 June; accepted 7 August 1986.

- Chang, C. & Cheng, H. *Sci. Sinica* **16**, 257–265 (1973).
- Chang, C. & Pan, Y. in *Geological and Ecological studies of the Qinghai-Xizang Plateau* Vol. 1, 1–18 (Science Press, Beijing, 1981).
- Bally, A. W. et al. *US Geol. Surv. Open File Rep.* **80-501** (1980).
- Sengor, A. M. C. in *Geological and Ecological Studies of the Qinghai-Xizang Plateau* Vol. 1, 51–56 (Science Press, Beijing, 1981).
- Tapponnier, P. et al. *Earth planet. Sci. Lett.* **52**, 355–371 (1981).
- Sengor, A. M. C. *Geol. Soc. Am. Spec. Pap.* **195**, (1984).
- Geological Institute of Qinghai-Xizang Plateau *Geological map of Qinghai-Xizang* 1:1,500,000, Chengdu (1980).
- Girardeau, J. et al. *Nature* **307**, 27–31 (1984).
- Girardeau, J. et al. *Geology* **13**, 330–333 (1985).
- Xu, R. et al. *J. Geol.* **93**, 41–57 (1985).
- Tapponnier, P. et al. *Nature* **294**, 405–410 (1981).
- Allegre, C. J. et al. *Nature* **307**, 17–22 (1984).
- Bassoulet, J. P. et al. in *Mission Franco-Chinoise au Tibet* 133–153 (CNRS, Paris, 1984).
- Yin, J. et al. in *A Scientific Guidebook to South Xizang (Tibet)* 17–104 (Acad. Sinica, Beijing, 1980).
- Jin, C. et al. *Kexue Tongbao* **29**, 1658–1661 (1984).
- Wan, Z. et al. *Tectonics of Yarlung-Zangbo suture zone, Xizang (Tibet)* (Geol. Bureau of Xizang, 1982).

- Burg, J.-P. et al. *Ecol. Geol. Helv.* **76**, 643–665 (1983).
- Maluski, H. et al. *Nature* **298**, 152–154 (1982).
- Xu, Z. *Oil Gas Geol.* **1**, 153–158 (1980).
- Dewey, J. F. & Burke, K. C. *J. Geol.* **81**, 683–692 (1973).
- England, P. & Houseman, G. *J. geophys. Res.* **91**, 3664–3676 (1986).
- Argand, E. *Proc. 13th Int. Geol. Cong.* **1(5)**, 171–372 (1924).
- Holmes, A. *Principles of Physical Geology* (Nelson, London, 1941).
- Carey, S. W. *Proc. R. Soc. Tasmania* **89**, 255–288 (1955).
- Powell, C. M. & Conaghan, P. J. *Earth planet. Sci. Lett.* **20**, 1–12 (1973).
- Barazangi, M. & Ni, J. *Geology* **10**, 179–185 (1982).
- Molnar, P. & Chen, W. *Nature* **273**, 218–220 (1978).
- Pozzi, J.-P. et al. *Nature* **297**, 319–321 (1982).
- Achache, J. et al. *J. geophys. Res.* **89**, 10311–10339 (1984).
- Zheng, S. *Palaeontology of Xizang* Vol. 1, 33–34 (1980).
- Liu, L. *Geology of the Qinghai-Xizang Plateau* **4**, 136–150 (1982).
- Molnar, P. & Tapponnier, P. *J. geophys. Res.* **83**, 5361–5384 (1978).
- Tapponnier, P. et al. in *Geological and Ecological Studies of the Qinghai-Xizang Plateau* Vol. 1, 43–40 (Science Press, Beijing, 1981).
- Tapponnier, P. et al. *Nature* **294**, 410–414 (1981).
- Burke, K. & Lucas, L. in *Tectonic evolution of Tethyan Regions* (NATO ASI, in the press).
- Cui, Z. & Yang, B. *Northwest. Seismol. J.* **1**, 57–61 (1979).
- Li, L. & Jia, Y. *Northwest. Seismol. J.* **3**, 61–65 (1981).
- Wu, X. et al. *Contr. Geol. Qinghai-Xizang (Tibet) Plateau* **4**, 1–18 (1982).
- Qian, F. et al. *Contr. Geol. Qinghai-Xizang (Tibet) Plateau* **4**, 121–130 (1982).

Structure, sequence and expression of the hepatitis delta (δ) viral genome

Kang-Sheng Wang*, Qui-Lim Choo*, Amy J. Weiner*, Jing-Hsiung Ou†, Richard C. Najarian*, Richard M. Thayer*, Guy T. Mullenbach*, Katherine J. Denniston‡, John L. Gerin* & Michael Houghton*

* Chiron Corporation, 4560 Horton Street, Emeryville, California, 94608, USA

† Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA

‡ Division of Molecular Virology & Immunology, Georgetown University, 5640 Fishers Lane, Rockville, Maryland 20852, USA

Biochemical and electron microscopic data indicate that the human hepatitis δ viral agent contains a covalently closed circular and single-stranded RNA genome that has certain similarities with viroid-like agents from plants. The sequence of the viral genome (1,678 nucleotides) has been determined and an open reading frame within the complementary strand has been shown to encode an antigen that binds specifically to antisera from patients with chronic hepatitis δ viral infections.

THE delta (δ) hepatitis antigen (HDAg) was first recognized by Rizzetto and coworkers in 1977 as a novel nuclear antigen appearing in certain Italian carriers of the hepatitis B virus (HBV)¹. Carriers expressing this HDAg exhibit a greater incidence of severe chronic active hepatitis and cirrhosis and the antigen is implicated in a substantial number of cases of fulminant hepatitis²⁻⁴. Chimpanzee transmission studies⁵ indicate that a defective viral agent is associated with δ hepatitis and that this agent requires HBV or the woodchuck hepadna virus⁶ in order to replicate. The hepatitis δ viral (HDV) agent replicates efficiently and suppresses helper replication which can lead to substantially higher titres of HDV relative to the hepadna virus^{5,7}. HDV is now known to be endemic among the HBV carrier population in Italy⁸ and has also been implicated in hepatitis cases throughout many parts of the world⁹ where it occurs either as a result of a super-infection of the HBV carrier state or as an acute co-infection⁸. High risk groups include intravenous drug addicts and haemophiliacs who have received clotting factors derived from pooled blood donations⁸.

In infectious sera, HDV particles of about 36 nm in diameter have been distinguished from the 42 nm Dane particle and 22 nm surface antigen moieties derived from HBV¹⁰. These particles appear to consist of an exterior HBV surface antigen and lipid coat which encapsidates HDAg and a δ -associated RNA molecule (HDV RNA)^{11,12}. This RNA molecule has been estimated by gel electrophoresis to contain about 1,700 nucleotides and is proposed to be single stranded because of its ribonuclease sensitivity¹³. Recently, a small (164 base pairs) cDNA clone (pKD3) was obtained in one of our laboratories and shown to hybridize specifically with HDV RNA derived from a variety of infectious sera¹⁴. HDV RNA is presumed to be the viral genome because of its association with viral preparations and the correlation between infectivity titre and HDV RNA concentration in an acute phase chimpanzee serum^{7,8}. We now report an extensive characterization of this HDV RNA.

Structure of RNA

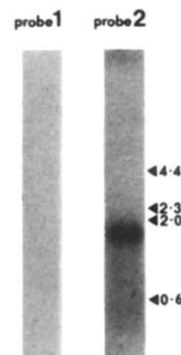
Since it has been suggested that HDV RNA lacks a poly(rA) segment¹³, we polyadenylated the RNA *in vitro* before synthesizing and cloning cDNA using the Okayama/Berg vector¹⁵. Recombinants were screened using an oligonucleotide hybridization probe derived from the sequence of pKD3. Two clones $\delta 1$ and $\delta 2$ were obtained containing 567 and 250 base pairs (bp) of cDNA sequence, respectively (excluding 'tails'). Strand-specific cDNA probes derived from clone $\delta 1$ were hybridized to the original HDV template RNA. Figure 1 demonstrates that only one of these probes hybridized to the HDV RNA thus

indicating that this RNA is single-stranded. The hybridizing RNA species consists of a closely migrating RNA doublet of ~1,700 nucleotides in length.

To isolate additional clones, a cDNA library generated by random priming HDV RNA was screened with a hybridization probe derived from clone $\delta 1$. Positive clones were then used as hybridization probes to screen for new, overlapping clones in the cDNA library. In this way independent clones $\delta 3b$, $\delta 4$, $\delta 7a$,

Fig. 1 Hybridization analysis of HDV RNA. Strand-specific hybridization probes 1 and 2 were derived from cDNA clone $\delta 1$ and hybridized individually to HDV RNA that had been previously electrophoresed through a denaturing RNA gel. The parallel migration of phage λ /HindIII restriction fragments (expressed in kbp) is indicated by arrows.

Methods. Strand-specific hybridization probes were prepared by cloning either a ~950 bp *PvuII*/HindIII restriction fragment (probe 1) or a ~450 bp *PvuII*/PstI fragment (probe 2) of clone $\delta 1$ into phage M13 vectors³⁹ to generate complementary single-stranded HDV templates. Hybridization probes were prepared by mixing 0.8 μ g of each template DNA with 0.1 μ g of hybridization probe primer (New England Biolabs) in 200 mM NaCl and incubating for 15 min at 37 °C after denaturing in a boiling water bath for 1 min. The annealed mixture was incubated at 15 °C for 2 h in 200 μ l containing 50 mM Tris/Cl pH 7.5, 5 mM MgCl₂, 10 mM β -mercaptoethanol, 50 μ g/ml BSA, 0.1 mM dATP, dGTP and dTTP, 14 μ M dCTP (1,000 C_i mmol⁻¹) and 250 units ml⁻¹ Klenow *E. coli* DNA polymerase 1 prior to phenol/chloroform extraction and chromatography on a G50(M) Sephadex (Pharmacia) column (20 cm. $\times$ 0.7 cm diameter) run in 50 mM NaCl, 0.1% SDS. Probe eluting in the void volume was hybridized to a 'Northern blot'⁴⁰ of a 1% agarose gel containing 2.2 M formaldehyde⁴¹ through which 10 ng HDV RNA and λ /HindIII restriction fragments had been electrophoresed. HDV RNA was derived from the serum of a chimpanzee during acute HDV infection and the preparation procedure included incubation with an excess of proteinase K and phenol/chloroform extraction^{14,42,43}. The blot was pre-hybridized at 42 °C overnight in 5 $\times$ SSC (1 $\times$ SSC is 150 mM NaCl, 15 mM sodium citrate), 50% (v/v) formamide, 50 mM Na₂HPO₄ pH 6.5, 0.2% (w/v) SDS, 2 $\times$ Denhardt's solution (50 $\times$ is 1% (w/v) BSA, 1% (w/v) Ficoll-400, 1% (w/v) polyvinylpyrrolidone) and 250 μ g ml⁻¹ salmon testis DNA (sonicated and denatured) prior to hybridizing at 42 °C overnight in 5 $\times$ SSC, 50% (v/v) formamide, 20 mM Na₂HPO₄ pH 6.5, 0.1% SDS, 1 $\times$ Denhardt's, 10% (w/v) dextran sulphate, 100 μ g ml⁻¹ salmon testis DNA and 10⁶ c.p.m. ml⁻¹ of probe. Filters were washed at 65 °C in 0.1 $\times$ SSC, 0.1% SDS and autoradiographed at -80 °C with an intensifier screen.



§ To whom correspondence should be addressed.

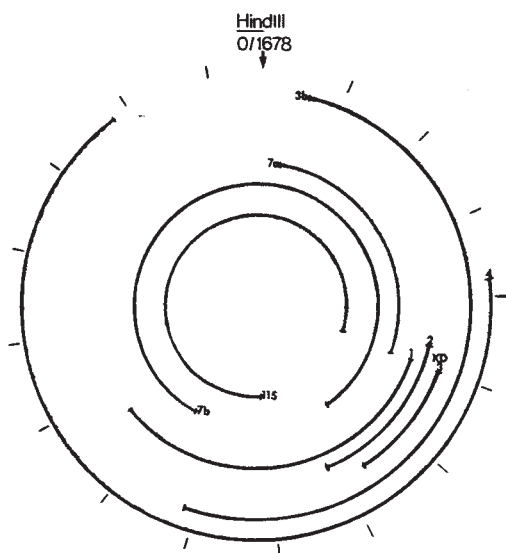


Fig. 2 Schematic alignment of the HDV cDNA sequences. The sequences of 7 independent HDV cDNA clones were determined (see Fig. 4) and aligned together with the initial clone pKD3. The scheme indicates the putative circular template RNA consisting of 1,678 nucleotides. The circle is shown as beginning in the centre of a unique *Hind*III restriction site. Increments of 100 nucleotides are indicated on the circumference.

$\delta 7b$ and $\delta 115$ were identified and shown to contain cDNA inserts of 829, 1,123, 474, 1,378 and 1,362 bp, respectively (excluding G-C 'tails').

Surprisingly, comparison of the nucleotide sequences of the cloned cDNAs suggested that the template RNA for at least some of them was a circular molecule of 1,678 nucleotides (Fig. 2). A linear model would require a molecule comprising ~2,180 nucleotides and containing terminal direct repeats of ~510 nucleotides. However, the HDV RNA migrates faster than 18S ribosomal RNA (1,869 nucleotides) on denaturing formaldehyde gels (ref. 14 and K.S.W., unpublished data) indicating a size of around 1,700 nucleotides. A hypothesis involving heterogeneous, permuted linear HDV RNA molecules could be consistent with the cloning and sequence data. A significant level of base substitution was observed in the sequences of overlapping regions of different clones but no insertion or deletion events were found, suggesting that the templates used to generate these seven clones are of uniform size.

The batch of HDV RNA used for cDNA cloning was examined in the electron microscope under completely denaturing conditions¹⁶. Circular and linear molecules were present in approximately similar proportions (see Fig. 3, *a-d*). The mean sizes of circles and linears observed in the electron microscope under these conditions were calculated to be 1,715 and 1,487 nucleotides respectively, values close to that of 1,678 nucleotides from sequence analysis and ~1,700 nucleotides from gel electrophoresis. The estimated smaller size of the linear molecules seen in the electron microscope probably reflects the presence of some degraded molecules, whereas circles were presumably all of full length.

Hybridization analysis of HDV RNA using probes derived from the random-primed cDNA clones yielded results similar to those shown in Fig. 1 (K.S.W., unpublished data) indicating that HDV RNA is completely single-stranded. It seems likely therefore that HDV contains a single-stranded, circular RNA molecule of 1,678 nucleotides, presumably covalently closed since cDNA synthesis appears to proceed around the molecule (Fig. 2) and circular forms are visible in the electron microscope under completely denaturing conditions for RNA (Fig. 3, *a-d*).

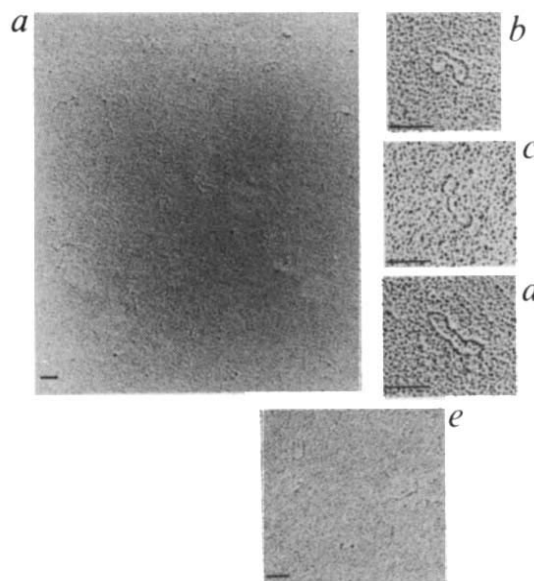


Fig. 3 Electron microscopy of HDV RNA. Panels *a-d*, completely denaturing conditions, *e*, weakly denaturing conditions. Several hundred molecules were observed under completely denaturing conditions. Circular and linear molecules were evident in similar proportions. Three RNA molecules of known length (777, 1,997 and 3,346 nucleotides) were used as standards under identical denaturing conditions and the length of these circles and linears was found to be 1,715 nucleotides (standard deviation of 175 from measurement of 22 different molecules) and 1,487 nucleotides (standard deviation of 290 from measurement of 31 different molecules) respectively. *a*, Typical field *b-d*, different, individual circular molecules. Several hundred molecules were observed under weakly denaturing conditions. All appeared as linear structures with an estimated size of 910 nucleotides (standard deviation of 70 from measurement of 38 different molecules). *e*, Typical field. Bar, 100 nm.

Methods. *a-d*, HDV RNA (50 ng) was heated at 70 °C for 15 min in 50 μ l of 90% (v/v) formamide, 10 mM Tris/Cl, pH 8.2, 1 mM EDTA, then 1.5 μ l of cytochrome *c* (1 mg ml⁻¹) was added. An aliquot (30 μ l) was then immediately spread onto the hypophase of 60% (v/v) formamide, 1.5 mM Tris/Cl, pH 8.2, 0.15 mM EDTA, pre-warmed to 70 °C. 30 s after spreading (at which time the temperature was >50 °C), the surface film formed by the sample was picked up on a carbon-coated grid. The grid was stained (45 s) in 0.002% (w/v) uranyl acetate in ethanol before shadowing with platinum and viewing in a Phillips 400 electron microscope¹⁶. *e*, HDV RNA (50 ng) was suspended at room temperature in 50 μ l of 40% (v/v) formamide, 10 mM Tris/Cl, pH 8.2, 1 mM EDTA and 60 μ g ml⁻¹ cytochrome *c* and spread onto the hypophase of 10% (v/v) formamide, 1 mM Tris/Cl pH 8.2, 0.1 mM EDTA and viewed in a Phillips 400 electron microscope⁴⁴.

When electron microscopy of this RNA was conducted under much less denaturing conditions only linear molecules were seen (Fig. 3, panel *e*), of mean size ~910 nucleotides.

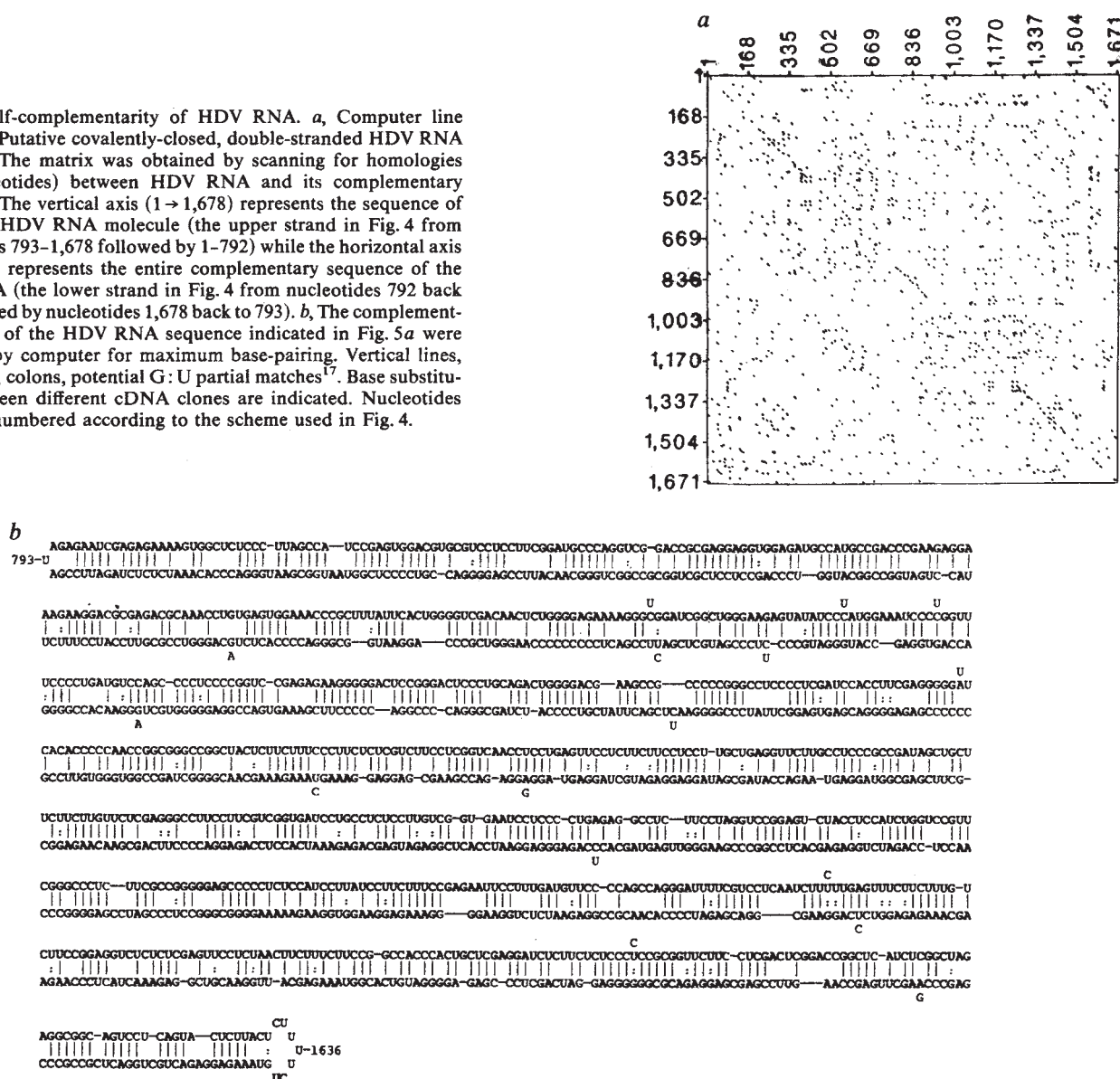
Nucleotide sequence

Figure 4 shows the nucleotide sequence of HDV RNA as deduced from analysis of the cDNAs in clones $\delta 1$, $\delta 2$, $\delta 3b$, $\delta 4$, $\delta 7a$, $\delta 7b$ and $\delta 115$. The region corresponding to the short cDNA sequence in clone pKD3 was derived from three independent clones generated in this study and differs in several positions from the sequence of pKD3¹⁴.

A prominent feature of the HDV RNA sequence is the high G+C content (60%) which results in the formation of stable secondary structure. Computer analysis of the sequence indicates an inherent ability of HDV RNA to self-anneal. Figure 5a shows a computer line matrix analysis where the entire HDV RNA sequence is scanned for homologies of 6 nucleotides or

Methods. Clones 81 and 82: the RNA (150 ng) was polyadenylated by pre-heating at 70 °C for 2 min in 50 mM Tris/Cl pH 7.9, 2 mM EDTA and incubating at 37 °C for 2 min in 50 mM Tris/Cl pH 7.9, 10 mM MgCl₂, 2.5 mM MnCl₂, 250 mM NaCl, 40 µM rATP, 2 mM DTT, 1,000 units ml⁻¹ RNAase inhibitor (Promega) and 100 units ml⁻¹ *Escherichia coli* poly(rA) polymerase (BRL) (50 µl). The RNA was extracted with phenol/chloroform and precipitated overnight at -20 °C with 2.5 volumes of absolute ethanol, recovered by centrifugation and resuspended in H₂O, then cloned into the Okayama/Berg vector¹⁵ using *E. coli* MC1061⁴⁵ as host. Ampicillin-resistant transformants were colony-hybridized⁴⁶ using an oligonucleotide (5' GAGTCGGAATCGAGCATCGGGAAGGGCATC) as hybridization probe (derived from the sequence of pKD3¹⁴). The probe was labelled by incubating 20 pmol at 37 °C for 1 h in 40 µl of 0.1 mM EDTA, 50 mM Tris Cl pH 9.0, 10 mM MgCl₂, 0.1 mM spermidine, 1 mM DTT containing 200 µCi [³²P]ATP(>5,000 Ci mmol⁻¹) and 700 units ml⁻¹ T4 polynucleotide kinase. After extraction with phenol and chloroform, the aqueous phase was bound to a Sep-Pak C18 cartridge (Millipore), eluted with 50% (v/v) CH₃OH, 50 mM CH₃COONH₄, pH 7.5, dried and resuspended in H₂O. The probe was hybridized at ~10⁶ c.p.m. ml⁻¹ using the hybridization and washing conditions described in Fig. 1. Positive clones 81 and 82 were thus identified. HDV clones 3b, 4, 7a, 7b and 115: HDV RNA (450 ng) was denatured in 10 mM CH₃GhOH at room temperature for 10 min and single strand cDNA was synthesized in a 50 µl reaction containing 50 mM Tris Cl pH 8.3, 10 mM MgCl₂, 70 mM KCl, 0.5 mM dATP, dGTP and TTP, 0.2 mM [³²P] dCTP (15 Ci mmol⁻¹), 600 units ml⁻¹ RNAase inhibitor (Promega), 14 mM β-mercaptoethanol, 5 mM DTT, 100 units ml⁻¹ AMV reverse transcriptase (Life Sciences), 50 µg ml⁻¹ actinomycin D, 0.01% (v/v) Triton X-100 and 4 µg ml⁻¹ random primers derived from a limit DNAase I digest of calf thymus DNA⁴⁷. After incubation at 37 °C for 1.5 h, the cDNA/RNA hybrids were extracted with phenol/chloroform, 5 µg chimpanzee liver RNA was added as carrier and the hybrids were precipitated with ethanol. Subsequent cloning steps have been described in detail elsewhere⁴⁸. Briefly, RNA was degraded (0.3 N NaOH, 1 h, 50 °C) and double stranded cDNA was prepared by allowing self-priming with reverse transcriptase. Hairpins were cleaved with S₁ nuclease and oligo(dC) tails were added using terminal transferase and annealed with G-tailed, PstI-cut pBR322 (New England Nuclear) and transformed into *E. coli* MC1061⁴⁵. Ampicillin-sensitive, tetracycline-resistant recombinants were colony-hybridized using a 'nick translated' 435 bp *Nco*I cDNA restriction fragment of clone 81 as hybridization probe. The cDNA inserts of positive clones (84 and 8115) were used as hybridization probes for the same library, giving clones 3b, 4, 7a, 7b and 115. Following restriction mapping, various restriction fragments of the cDNA inserts from each clone were sub-cloned into M13 vectors (mp18 and mp19)³⁹ and sequenced using the dideoxy chain termination method⁴⁹. The sequence shown was deduced from complementary strands of at least two overlapping clones. Areas of uncertainty were confirmed using the techniques of Maxam and Gilbert⁵⁰.

Fig. 5 Self-complementarity of HDV RNA. *a*, Computer line matrix, *b*, Putative covalently-closed, double-stranded HDV RNA structure. The matrix was obtained by scanning for homologies (≥ 6 nucleotides) between HDV RNA and its complementary sequence. The vertical axis (1 $\rightarrow$ 1,678) represents the sequence of the entire HDV RNA molecule (the upper strand in Fig. 4 from nucleotides 793–1,678 followed by 1–792) while the horizontal axis (1 $\rightarrow$ 1,678) represents the entire complementary sequence of the HDV RNA (the lower strand in Fig. 4 from nucleotides 792 back to 1 followed by nucleotides 1,678 back to 793). *b*, The complementing halves of the HDV RNA sequence indicated in Fig. 5*a* were arranged by computer for maximum base-pairing. Vertical lines, base-pairs; colons, potential G:U partial matches¹⁷. Base substitutions between different cDNA clones are indicated. Nucleotides are numbered according to the scheme used in Fig. 4.



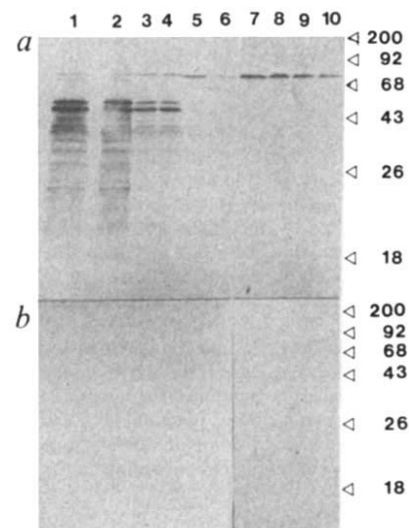
more with its complementary sequence. Homology is indicated by a short line of length depending on the size of the homologous region. This analysis indicates regions of intra-molecular and self-complementarity, and putative self-annealing within the HDV RNA molecule itself. Such regions extend throughout the entire HDV RNA sequence as can be seen from the semi-continuous diagonal line that begins at the origin and extends throughout the matrix figure. Figure 5*a* indicates that nucleotides 793–1,632 (Fig. 4) of the HDV RNA molecule (*y* coordinates 1–839 of Fig. 5*a*) can extensively complement the other half of the HDV RNA sequence (*x* coordinates 1–839 of Fig. 5*a*). Using computer analysis to obtain maximum complementation in this region a highly base-paired, covalently-closed structural model for the HDV RNA molecule can be proposed (Fig. 5*b*). Sixty-one per cent of the nucleotides are base-paired in this putative rod-like structure and a further 6% can take part in G-U matches¹⁷. The linear molecules observed in the electron microscope under weak denaturing conditions (Fig. 3*e*) were approximately half the size of the circular molecules seen under completely denaturing conditions, consistent with their being double-stranded, rod-like structures.

The HDV RNA sequence and the complementary anti-HDV RNA strand contain numerous open reading frames (ORFs) the largest of which are indicated in Table 1. Many of these ORFs contain potential initiator ATG codons which, if responsible for translation initiation, could encode polypeptides of up to 215 amino acids before taking into account possibilities of splicing and/or frameshifting events.

In the comparison of the HDV RNA sequence with all other known viral and viral-like sequences (using computer analysis), the most significant homology detected was with the S strand of the ayw strain of HBV¹⁸. Allowing for many gaps in the alignment, a low level of homology was apparent which had some statistical significance ($P=0.02$) compared with the homologies obtained using randomized nucleotide sequences (data not shown). The greatest region of homology is in a small ORF in HDV RNA (nucleotides 1,477–1,678/1,678–2; Fig. 4) with a coding potential of 90 amino acids, although none of these is methionine or hence a possible translation initiation site. However, part of the potential polypeptide encoded by this ORF has some homology with the carboxy terminus of the HBV (ayw) polymerase^{18,19} with the 25 amino acids encoded by

Fig. 6 Immunological analysis of polypeptides expressed in HDV cDNA recombinant bacteria. ORFs 5 and 6 were expressed in the form of fusion polypeptides with the human enzyme superoxide dismutase (SOD) under the control of the inducible 'Tac' promoter²⁰, then reacted with either *a*, antiserum from a patient chronically infected with HDV or *b*, a pool of control antisera from previous infections with either hepatitis A, B or Non-A, Non-B viral agents. The bacteria were transformed with: lanes 1 and 2, pSOD-ORF 5 (induced with IPTG); lanes 3 and 4, pSOD-ORF 5 (no IPTG); lane 5, parent pSOD vector (induced with IPTG); lane 6, parent pSOD vector (no IPTG); lanes 7 and 8, pSOD-ORF 6 (induced with IPTG); lanes 9 and 10, pSOD-ORF 6 (no IPTG). The migration of protein standard (expressed as r.m.m. $\times 10^{-3}$) is indicated.

Methods. HDV cDNA restriction fragments corresponding to the regions of ORF 5 and ORF 6 were inserted into the region of plasmid pSOD16cf2²⁰ that specifies the carboxy terminus of SOD, allowing the synthesis of fusion proteins. To make the ORF 5 fusion protein, pSOD16cf2 was cleaved with *Nco*I, 'blunt-ended' with the Klenow fragment of *E. coli* DNA polymerase I and digested with *Sall* prior to extracting the large *Nco*I/*Sall* vector fragment from agarose gels following electrophoresis. Clone δ 115 was digested with *Sst*II, 'blunt-ended' and restricted with *Sall*; the fragment (~600 bp) was recovered from agarose gels. The fragments were ligated using T4 DNA ligase and transformed into *E. coli* strain D1210³¹. Recombinants were identified by restriction mapping and sequence determination of the junction between the human SOD cDNA and the HDV cDNA. Recombinant bacteria expressing ORF 6 fusion proteins were made in a similar way using a *Sma*I/*Eco*RI digested vector fragment and a 622 bp *Sma*I/*Eco*RI fragment derived from clone δ 4. Overnight cultures of recombinant bacteria were diluted 100 $\times$ into L-broth containing 100 μ g ml⁻¹ ampicillin and grown at 37 °C to an OD₆₅₀ of 0.6 before either extraction or addition of 1 mM IPTG (Sigma) for 4 h to maximize expression before extraction. Bacteria were pelleted, resuspended in 10% (v/v) glycerol, 50 mM DTT, 3% (w/v) SDS, 60 mM Tris/Cl⁻ pH 6.8 and freeze thawed (3 cycles). After incubation in a boiling water bath for 5 min, samples were electrophoresed through 12% (w/v) polyacrylamide gels containing SDS³² and electroblotted onto nitrocellulose filters⁵³. The filters were pre-incubated with 5% (v/v) goat serum, then incubated with a 1:300 dilution of antisera in phosphate-buffered saline (PBS) containing 0.3% Tween 20 (Sigma) and 5% (v/v) goat serum for 1 h at room temperature. The filters were then incubated with a 1:200 dilution of horse radish peroxidase conjugated to goat anti-human IgG in the presence of the chromogen 4-chloro-1-naphthol (Biorad).



nucleotides 1,483–1,557 showing 52% homology with a carboxy-terminal region of HBV polymerase (Fig. 4, legend). Although there are a number of direct repeats and palindrome-like sequences in HDV RNA (Fig. 4) of which some may be important in replication, the direct repeats DR1 and DR2¹⁹ that play an important role in replication of hepadna viruses are not present.

Expression of HDV

In an attempt to identify the ORF(s) coding for HDAg (previously termed δ antigen), two long ORFs (ORFs 5 and 6; Fig. 4 and Table 1) apparent in the sequence complementary to HDV RNA were selected for study. These ORFs were sub-cloned into a bacterial plasmid which uses the strong inducible 'Tac' promoter to express the ORFs as a fusion protein containing the human superoxide dismutase polypeptide at the amino terminus and the HDV-encoded polypeptides at the carboxy terminus²⁰. Fusion proteins synthesised in this way were transferred to nitrocellulose after gel electrophoresis and then monitored for specific binding to an antiserum derived from a patient with a chronic HDV infection. Figure 6*a* shows that ORF 5 but not ORF 6 fusion polypeptides specifically bound HDV antiserum. A different batch of human HDV antiserum also binds the ORF 5 fusion proteins (Q.L.C., A.J.W., unpublished observation) whereas control antisera did not (Fig. 6*b*). Both HDV antisera have been shown to contain antibodies capable of binding to HDV viral polypeptides²¹.

The estimated size of the largest major immunoreactive ORF 5 polypeptide observed in Fig. 6*a* (49,000 daltons) is consistent with a fusion polypeptide containing 154 amino acids of superoxide dismutase and 205 amino acids specified by ORF 5. Translation of ORF 5 in the other two reading frames (or in the opposite orientation) would produce much smaller polypeptides because of the presence of stop codons. The size heterogeneity observed with the ORF 5 fusion products may be a result of bacterial proteolysis. The failure to detect specific immunoreactive ORF 6 fusion polypeptides was not due to poor bacterial expression since, when monitored for binding to rabbit antiserum raised against human superoxide dismutase, both ORF 5 and ORF 6 fusion polypeptides were present at similar levels (A.J.W., unpublished data). Thus ORF 5 contributes to the syn-

thesis of an immunogenic HDV polypeptide, which has unusual structural features. Clusters of basic and acidic residues appear throughout most of the molecule, except for the carboxy terminus which is much less charged and more hydrophobic in character. With an overall positive charge, the molecule could bind nucleic acid, as expected for the HDAg antigen, which is found inside the viral particle with the viral RNA. The hydrophobic carboxy terminus may interact with the coat of the HDV particle, which consists of HBV surface antigens and host-derived lipid components^{11,12}. There is also a potential *N*-linked glycosylation site²² in the putative ORF 5 polypeptide (Fig. 4).

Conclusion

Our evidence suggests that HDV contains a single-stranded covalently-closed circular RNA molecule of 1678 nucleotides which can form extensive secondary structure. The observation that HDV cDNA encodes an antigen that specifically binds antisera from patients with chronic HDV infections confirms previous data suggesting the genomic character of HDV RNA.

The unusual properties of the HDV genome resemble those of a number of different viroid-like molecules that cause a variety of diseases in plants. Viroids^{23,17} are small single-stranded circular RNA molecules of high G-C content that form double-stranded rod-like structures²⁴ which are not encapsidated and do not require a helper agent for replication. Virusoids, a class of viroid-like molecules, are one of the two essential components of the bipartite RNA genomes of recently discovered plant viruses²⁵. The other component is a large linear RNA molecule that may encode the replication function for both molecules. Finally, the satellite of the Lucerne Transient Streak virus²⁶, previously considered a tomato virusoid²⁷, has a viroid-like structure and, as noted earlier²⁸, is similar to HDV in requiring a largely unrelated helper virus (which it inhibits) for replication.

The sequence GAAAC occurs in all of these viroid-like molecules (except the hop stunt viroid²⁹) and in the HDV genome (nucleotides 88–92, Fig. 4). The sequence GAUUUU also exists in the genomes of virusoids and HDV (nucleotides 1453–1458, Fig. 4). In addition, the HDV genome displays some homology (nucleotide regions 75–94 and 552–571 (Fig. 4) both

Table 1 Large open reading frames (ORFs) within HDV RNA and the complementary anti-HDV RNA strand

Strand	ORF number	Translational frame	Nucleotide* position	Total no. of amino acids	No. of amino acids beginning with first methionine
HDV RNA	1	2	539	165	152
HDV RNA	2	3	786	169	134
HDV RNA	3	2	1,607	120	0
HDV RNA	4	3	1,296	115	68
Anti-HDV RNA	5	1	1,618	222	215
Anti-HDV RNA	6	2	1,140	221	179/34†
Anti-HDV RNA	7	3	506	148/80	148/80‡
Anti-HDV RNA	8	3	1,340	109	0
Anti-HDV RNA	9	1	91	101	0

Only ORFs longer than 300 nucleotides are shown.

* The position of the first nucleotide in each open reading frame is indicated according to the numbering shown in Fig. 4.

† Ambiguity arising from clonal heterogeneity at position 1012 (Fig. 4).

‡ Ambiguity arising from clonal heterogeneity at position 264 (Fig. 4).

display 60–65% homology) with a viroid central conserved region, possibly involved in the processing of replication intermediates³⁰. The functional significance of these homologies and of the general similarities with plant agents remains to be established. It is possible that HDV replicates in a 'rolling-circle' mechanism through the action of a replicase activity, similar to that suggested for viroids³¹.

Although the sequences of the numerous HDV cDNA clones give no indication of size heterogeneity in the original HDV template RNA, two closely-migrating RNA species were observed following gel electrophoresis under denaturing conditions (Fig. 1). The species are highly homologous as hybridization probes remained annealed to the doublet after very stringent washing (Fig. 1). It is possible that as has been shown for plant viroids^{16,29,32}, the two species represent circular and 'nicked' linear forms of the same molecule. This hypothesis is supported by the observation of circles and linear molecules of similar size in the electron microscope under completely denaturing conditions (Fig. 3a) and the generation of clones $\delta 1$ and $\delta 2$ following an initial polyadenylation of HDV RNA termini.

It should also be noted that in the electron microscope a few circular molecules appear to possess a tail (Fig. 3d). Although this is probably an artefact of the cytochrome C background, the possibility that a subset of HDV RNA molecules have circular lariat structures similar to those previously described for certain processed introns^{33,34} cannot be excluded.

A low level of homology is apparent between the HDV genome and the S strand of HBV (strain ayw). The implications of this are unclear as yet in the absence of further information about the replication of HDV. The degree of homology is consistent with our recent work showing that the genomes of HDV and HBV do not cross-hybridize¹³ even under low stringency conditions³⁵.

There are many ORFs in both the genomic and anti-genomic strands of HDV. Interestingly, an ORF in the anti-genomic strand (ORF 5; Fig. 4 and Table 1) appears to encode an immunogenic viral antigen that is specifically recognized by antisera derived from patients with chronic HDV infections, whereas another ORF in the same anti-genomic strand (ORF 6, Table 1) does not. A similar negative result has also been obtained with an ORF (ORF 1, Table 1) in the genomic strand (A.J.W., unpublished data). Combined with the fact that eukaryote ribosomes cannot translate circular RNA molecules^{36,37} and the observed inability of an *in vitro* rabbit reticulocyte lysate to translate the HDV genome (K.S.W., unpublished data), our data suggest that HDV is a negative-stranded virus. However, further studies are necessary before we can rule out any coding function for the genomic strand.

If translation of ORF 5 initiates at the first methionine codon and terminates at the first in-frame stop codon a polypeptide of 215 amino acids (about 25,000 daltons) would be synthesized (Fig. 4). Although earlier work using gel filtration analysis indicated that the molecular weight of HDAg was about 68,000³⁸, very recent analyses employing denaturing electrophoresis and Western blot techniques indicate that the HDV particle contains antigens of around 24,000 and 27,000 daltons²¹ that bind specifically to the antisera used in this study. This data is therefore consistent with ORF 5 in the anti-genomic strand being responsible for the synthesis of at least one of these HDV polypeptides.

We thank S. Brown-Shimer for sequencing work, G. Bell, P. Luciw, R. Sanchez-Pescador, P. Valenzuela and M. Power for critical evaluation of the manuscript and assistance with computer analysis, B. Baroudy for helpful discussion, Toni H. Jones for typing the manuscript, and L. Overby for advice and encouragement. This work was funded by Chiron Research Laboratories, Emeryville, California.

Received 6 June; accepted 20 August 1986.

1. Rizzetto, M. *et al.* *Gut* **18**, 997–1003 (1977).
2. Jacobson, I. M. *et al.* *Hepatology* **5**, 188–191 (1985).
3. Govindarajan, S., Chin, K. P., Redeker, A. G. & Peters, R. L. *Gastroenterology* **86**, 1417–1420 (1984).
4. Arico, S. *et al.* *Lancet* **ii**, 356–358 (1985).
5. Rizzetto, M. *et al.* *J. infect. Dis.* **141**, 590–602 (1980).
6. Ponzetto, A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **81**, 2208–2212 (1984).
7. Ponzetto, A. *et al.* *J. infect. Dis.* (in the press).
8. Rizzetto, M. *Hepatology* **3**, 729–737 (1983).
9. Purcell, R. H. & Gerin, J. L. in *Viral Hepatitis and Delta Infection* (ed. Verme, G., Bonino, F. & Rizzetto, M.) 113–119 (Liss, New York, 1983).
10. Rizzetto, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **77**, 6124–6128 (1980).
11. Bonino, F. *et al.* *Hepatology* **1**, 127–131 (1981).
12. Bonino, F. *et al.* *Infect. Immun.* **43**, 1000–1005 (1984).
13. Hoyer, B. *et al.* in *Viral Hepatitis and Delta Infection* (ed. Verme, G., Bonino, F. & Rizzetto, M.) 91–97 (Liss, New York, 1983).
14. Denniston, K. J. *et al.* *Science* **232**, 873–875 (1986).
15. Okayama, H. & Berg, P. *Molec. cell. Biol.* **2**, 161–170 (1982).
16. Randles, J. W. & Hatta, T. *Virology* **96**, 47–53 (1979).
17. Haseloff, J., Mohamed, N. A. & Symons, R. H. *Nature* **299**, 316–321 (1982).
18. Galibert, F., Mandant, E., Fitoussi, F., Tiollais, P. & Charnay, P. *Nature* **281**, 646–650 (1979).
19. Tiollais, P., Pourcel, C. & Dejean, A. *Nature* **317**, 489–495 (1985).
20. Steimer, K. S. *et al.* *J. Virol.* **58**, 9–16 (1986).
21. Bergmann, K. F. & Gerin, J. L. *J. infect. Dis.* (in the press).
22. Walter, P., Gilmore, R. & Blobel, G. *Cell* **38**, 5–8 (1984).
23. Diener, T. O. *Adv. Virus Res.* **28**, 241–283 (1983).
24. Riesner, D. *et al.* *Biophys. struct. Mech.* **9**, 145–170 (1983).
25. Haseloff, J. & Symons, R. H. *Nucleic Acids Res.* **10**, 3681–3691 (1982).
26. Jones, A. T., Mayo, M. A. & Duncan, G. H. *J. gen. Virol.* **64**, 1167–1173 (1983).
27. Keese, P., Bruening, G. & Symons, R. H. *FEBS Lett.* **159**, 185–190 (1983).
28. Kaper, J. M. & Tousignant, M. F. *Endavour, New Ser.* **8**, 194–200 (1984).
29. Ohno, T., Takamatsu, N., Meshi, T. & Okada, Y. *Nucleic Acids Res.* **11**, 6185–6197 (1983).
30. Diener, T. O. *Proc. natn. Acad. Sci. U.S.A.* **83**, 58–62 (1986).
31. Branch, A. D. & Robertson, H. D. *Science* **223**, 450–455 (1984).
32. Hashimoto, J., Suzuki, K. & Uchida, T. *J. gen. Virol.* **66**, 1545–1551 (1985).
33. Halbreich, A., Pajot, P., Foucher, M., Grandchamp, C. & Slonimski, P. *Cell* **19**, 321–329 (1980).
34. Zeitlin, S. & Efstratiadis, A. *Cell* **39**, 589–602 (1984).

35. Weiner, A. J. *et al. J. med. Virol.* (in the press).
36. Kozak, M. *Nature* **280**, 82–85 (1979).
37. Konarska, M., Filipowicz, W., Domdey, H. & Gross, H. J. *Eur. J. Biochem.* **114**, 221–227 (1981).
38. Rizzetto, M., Shih, J. W.-K. & Gerin, J. L. *J. Immun.* **125**, 318–324 (1980).
39. Messing, J. *Meth. Enzym.* **101**, 20–37 (1983).
40. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
41. Goldberg, D. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5794–5798 (1980).
42. Ponzetto, A. *et al. Hepatology* (in the press).
43. Smedile, A. *et al. Hepatology* (submitted).
44. Davis, R. W., Simon, M. & Davidson, N. *Meth. Enzym.* **413–428** (1971).
45. Casabadan, M. & Cohen, S. *J. Molec. Biol.* **138**, 179–207 (1980).
46. Grunstein, M. & Hogness, D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965 (1975).
47. Taylor, J. M., Illmensee, R. & Summers, J. *Biochim. biophys. Acta* **442**, 324–330 (1976).
48. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning* 229–242 (Cold Spring Harbor Laboratory (New York), 1982).
49. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
50. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
51. Sadler, J. R. *et al. Gene* **8**, 279–300 (1980).
52. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
53. Towbin, H. T., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).

LETTERS TO NATURE

Double clusters and gravitational lenses

C. S. Crawford, A. C. Fabian & M. J. Rees

Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

Clusters of galaxies can act as gravitational lenses¹ which produce double quasars with splittings as large as 1–2 arc min. The potential well of the cluster must be deep enough on the scale of the image to focus the light, which means that the velocity dispersion within the cluster must be high and the core radius small. These requirements are reduced if there are two (or more) clusters along the line of sight. Paczynski and Gorski² have modelled the triple image produced by two clusters without pursuing the probability of such an alignment in any detail. Here we predict the number of double quasar images in the sky that are produced by double clusters. The clustering of clusters³ and the occurrence of binary clusters^{4,5} enhances the probability of double clusters well above that expected from a purely random spatial distribution. Several double images are expected. Double clusters should be considered for apparent double quasar images such as Hazard 1146 + 111 B,C (refs 6, 7) before cosmic strings⁸ or supermassive black holes⁹ are required. Although H1146 + 111B and C may not be due to a gravitational lens^{10,11}, the discovery of a *bona fide* wide double quasar image will be important in the study of the evolution of the clustering of clusters.

We have used two approximate methods to estimate the number of wide double images. The first works from the number density of rich clusters¹² and the clustering of clusters³. There is no strong richness/velocity dispersion relation^{13,14} but we shall assume that two clusters of Abell richness class 2 contribute an effective line-of-sight velocity dispersion, $\sigma_{||}$, of 2,000 km s⁻¹. The angular separation of the gravitational images is $\sim 2(\sigma_{||}/2,000 \text{ km s}^{-1})^2$ arc min. This requires that they each have $\delta_{||}$ of $\sim 1,400 \text{ km s}^{-1}$. The present number density of clusters of this richness and greater¹² is, $n(0) \approx 6 \times 10^{-8} \text{ Mpc}^{-3}$ (we adopt $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$). A similar result is obtained using the Piccinotti *et al.*¹⁵ X-ray sample and the Quintana-Melnick¹⁶ X-ray luminosity/velocity dispersion correlation. The fraction of the sky covering by double clusters aligned within an angle closer than the observed image separation θ (assuming no spatial correlation) is

$$f = \left(\frac{4c^3}{H^3} \pi \theta^2 n(0) \int_{\zeta_1}^{\zeta_2} (\zeta^{-3/2} + \zeta^{-5/2} - 2\zeta^{-2}) d\zeta \right)^2 \quad (1)$$

where $\zeta = 1 + z$; and ζ_1 to ζ_2 is the range of redshift considered for the clusters. The covering fraction f is enhanced by the clustering of clusters³ which, with scaling to $z \sim 0.5$, corresponds to a factor of $19/\theta$, where θ is in arc minutes. Then if $\theta \approx 2$ arc min we have an f of 10^{-5} . This implies that there would be about four double images in the sky if there are more than 10 appropriate background quasars per square degree. We have taken $z = 0.5$, $\zeta_1 = 1.25$ and $\zeta_2 = 1.75$ so the values are appropriate

to quasars of redshift ~ 1 . The velocity dispersion requirements are lessened if the clusters image the more numerous background quasars with redshifts of between 2 and 3. Including Abell richness class 1, we estimate an f of 5×10^{-5} . Consequently, there are likely to be ~ 200 wide quasar images in the sky if the background quasar density is as high as 100 per square degree.

Bahcall and Soneira³ remark that the clustering of clusters increases with richness so we may expect that it is most pronounced for the more massive objects required here. The above estimate involves a large extrapolation of their correlation function to small scales (it was measured on more nearby clusters with scales of degrees instead of arc minutes). There are, however, a reasonable fraction of nearby clusters observed to be double^{4,5} with the separation required by a 2 arc min lens image. An excellent candidate is the A399/401 cluster system which has velocity dispersions¹⁷ of 1,424 and 1,294 km s⁻¹, respectively (both with an uncertainty of $\sim 400 \text{ km s}^{-1}$). The apparent separation of these clusters at $z = 0.073$ is 33 arc min, which corresponds to 4.8 arc min at $z = 0.5$. Simple geometry then indicates that the probability of observing them with a separation of 2 arc min is ~ 0.1 . For our second method, we now assume that the A399/401 pair is not unique and estimate its spatial density to be approximately one per volume corresponding to $z = 0.073$. There are then ~ 70 pairs aligned to within 2 arc min over the sky out to $z = 0.75$. Again, we expect about four double quasar images in the sky from 10 background quasars per square degree.

The timescale over which double clusters may last with such small separations of a few megaparsecs is unknown, but it can take several crossing times (each $\sim 10^9$ yr) for the clusters to merge¹⁸. The Coma, CA0340–538 clusters and other binary clusters are large clusters containing two giant galaxies and could well have been A399/401-like pairs at redshifts $z \sim 0.5$. Indeed, the hierarchical clustering of clusters strongly suggests that double clusters were more common in the past.

We have ignored the possible brightness enhancement of the background quasars by the gravitational lens, which depends on the precise light paths. Any Sunyaev-Zeldovich¹⁹ effect due to hot gas in the clusters is a factor of two smaller in a double cluster than from a single very massive cluster (and possibly even less if the beam size is not matched to the cluster size). No single cluster has yet been reported with $\sigma_{||}$ of 2,000 km s⁻¹. Our estimates suggest that several quasars at $z \geq 1$ are gravitationally lensed into double images of 2 arc min separation by double clusters. The discovery of these images will enable the clustering of clusters to be pursued to distances well beyond that of the Abell catalogue.

Received 16 June; accepted 10 July 1986.

1. Narayan, R., Blandford, R. & Nityananda, R. *Nature* **310**, 112–115 (1984).
2. Paczynski, B. & Gorski, K. *Astrophys. J.* **248**, L101–L104 (1984).
3. Bahcall, N. A. & Soneira, R. M. *Astrophys. J.* **270**, 20–38 (1983).
4. Ulmer, M. P. & Cruddace, R. G. *Astrophys. J. Lett.* **246**, L99–L103 (1981).
5. Forman, W. *et al. Astrophys. J.* **243**, L133–L136 (1981).
6. Hazard, C., Arp, H. & Morton, D. C. *Nature* **282**, 271 (1979).
7. Turner, E. *et al. Nature* **321**, 142–144 (1986).
8. Gott, J. R. *Nature* **321**, 420–421 (1986).

9. Paczynski, B. *Nature* **321**, 419-420 (1986).
10. *Nature* **321**, 377 (1986).
11. Shaver, P. A. & Christiani, S. *Nature* **321**, 585-586 (1986).
12. Bahcall, N. A. *Astrophys. J.* **232**, 689-698 (1979).
13. Bahcall, N. A. *Astrophys. J.* **247**, 787-791 (1985).
14. Rivolo, A. R. *Nature* **317**, 788-789 (1985).
15. Piccinotti, G. *et al. Astrophys. J.* **253**, 485-503 (1982).
16. Quintana, H. & Melnick, J. *Astrophys. J.* **87**, 972-978 (1982).
17. Jones, C. & Forman, W. *Astrophys. J.* **276**, 38-55 (1984).
18. McGlynn, T. A. & Fabian, A. C. *Mon. Not. R. astr. Soc.* **208**, 709-718 (1984).
19. Sunyaev, R. A. & Zel'dovich, Ya. B. *Comm. Astr. Space Sci.* **4**, 173-179 (1972).

Multiple quasars for multiple images

John N. Bahcall, Neta A. Bahcall
& Donald P. Schneider

Institute for Advanced Study, Princeton, New Jersey 08540, USA and
Space Telescope Science Institute, Baltimore, Maryland 21218, USA

Three quasar pairs that have been identified as the products of gravitational lenses have no obvious visible lenses¹⁻³. They are: 1146+111 (redshift $z = 1.01$; pair separation $\Delta\theta = 157$ arc s)³, 1635+267 ($z = 1.96$; $\Delta\theta = 4$ arc s)¹ and 2345+007 ($z = 2.15$; $\Delta\theta = 7$ arc s)². An even number (two) of quasar images is observed in each case, although an odd number of images is produced by gravitational lenses with non-singular potentials⁴. The absence of a visible lens creates, in the lens hypothesis, a severe 'missing matter' problem. The missing lenses range in estimated mass^{3,5} from $>10^{12} M_{\odot}$ for 1635+267 (where $M_{\odot}$ is the mass of the Sun) to $\sim 10^{15} M_{\odot}$ for 1146+111. We investigate here the possibility that these quasar images are pairs of physically distinct quasars in galaxy associations. We show that, at the same redshifts as the quasars, only galaxy associations of exceptional richness would be identifiable in the existing data. Less prominent groups of galaxies, such as those known to be associated with nearby quasars⁶, would not be visible. We calculate the probability of pairs of quasars appearing in galaxy associations and find that physical pairs could appear with the observed frequency if, as expected, quasars are more common relative to galaxy associations at earlier epochs. We also discuss observations that can distinguish between the hypothesis of a gravitational lens and that of physical pairs.

Observations of nearby quasars support the suggestion that quasars may come in groups. Stockton⁶⁻⁸ has shown that at least one third of all nearby quasars are surrounded by small groups of galaxies at the same redshift. For later reference, we note that the observed quasar-galaxy separations range from ~ 10 to 400 kpc (for Hubble constant $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$)⁶⁻⁸; the upper limit is fixed by the size of the survey fields. The mean velocity difference between the quasar and the associated galaxies is consistent with zero and has a standard deviation of $\sim 400 \text{ km s}^{-1}$.

Figure 1a illustrates the difficulty of detecting associated galaxies around the high redshift quasar pairs if the galaxies are at the same redshift as the quasar. One has the best chance of observing associated galaxies for 1146+111 (the 'wide pair') as it has the lowest redshift. Figure 1a shows a deep image of the field obtained at the Palomar 5-m telescope with the 4-Shooter⁹. The limiting magnitude for galaxies on this *i*-band (8,000 Å) exposure is ~ 23.5 mag. No prominent association of galaxies is apparent in Fig. 1a.

Would galaxies associated with the quasars have been identified in this exposure? We answered this question by using data on low-redshift clusters of galaxies transformed to take account of *k*-corrections¹⁰ (no evolution) and of metric differences (for deceleration parameter $q_0 = 1/2$). We selected *i*-band exposures of 1146+111 because this filter is most likely to reveal any associated galaxies at $z \approx 1$. Figure 1b shows the simulated data when the cluster CL0024+1654, which is one of the richest

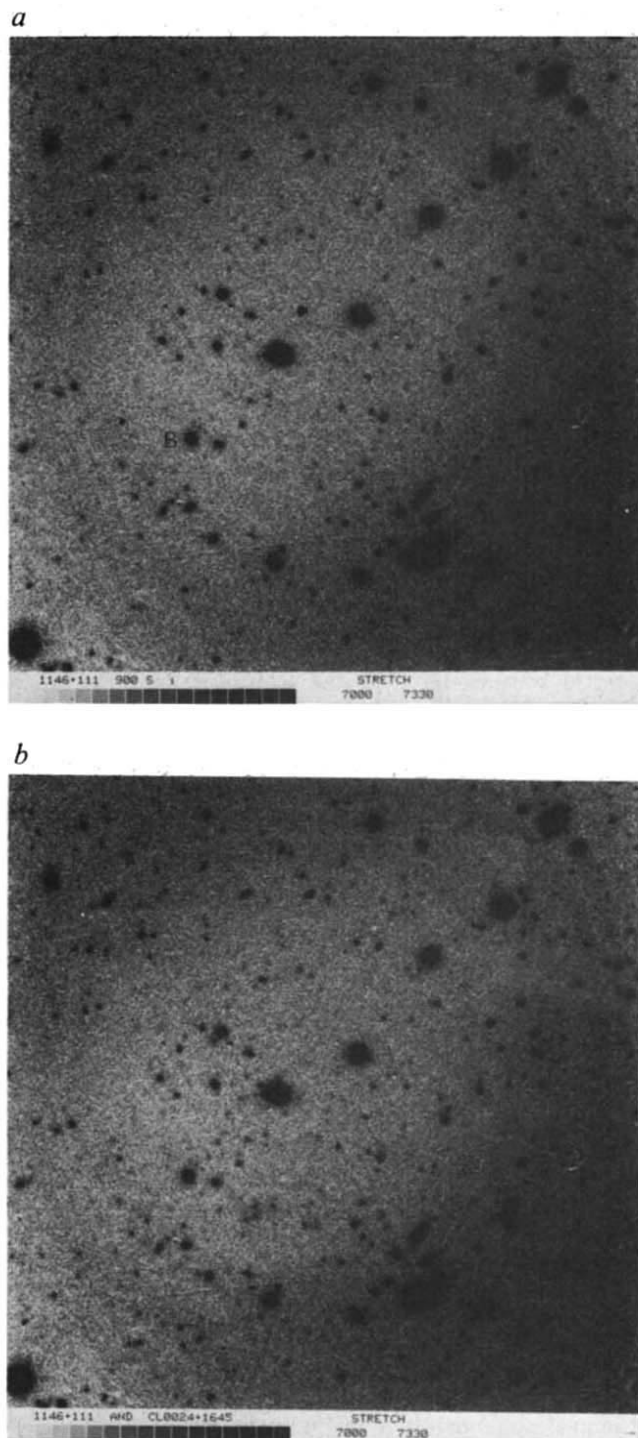


Fig. 1 *a*, An *i*-band CCD (charge-coupled device) frame of the field containing the quasar-pair 1146+111 B, C. North is up, east is to the left, and the picture is 250 arc s on a side. The seeing is 1.4 arc s (FWHM). *b*, As *a*, except that a frame of the very rich cluster of galaxies CL0024+1654 has been added to the quasar picture. The image of the cluster ($z = 0.39$) has been processed so that it appears as if it were at the redshift of the pair of quasars ($z = 1.01$). The centre of the cluster is ~ 30 arc s south-west of quasar B.

known galaxy clusters (Abell richness ≥ 4 (ref. 11)), is transformed to the redshift of the quasar pair and superimposed on the field. Only the brightest cluster galaxies are visible on the simulated image. We conclude that a cluster of Abell richness 4 would have been detected around 1146+111, but only barely. A similar simulation made with a typical richness-3 cluster shows

that such a system would at best be only marginally detected.

Galaxy associations at the same redshifts as the quasar images considered here would have escaped detection on state-of-the-art images if the associations were less rich than Abell class 3. Most (>95%) galaxy associations observed at low redshift satisfy this requirement.

Would we expect the existing observations to have revealed three pairs of quasars in galaxy associations if the quasars are randomly distributed among the associations? The number of pairs of quasars that would be expected is

$$\text{Number of quasar pairs} = f^2 \left(\frac{n_{\text{QSO}}}{n_{\text{assoc}}} \right) \frac{N_{\text{QSO}}}{2} \quad (1)$$

where f is the probability that a given quasar is found in a (galaxy) association. The number density of galaxy associations is n_{assoc} ; n_{QSO} is the quasar number density. For definiteness, we assume that the total number of relevant quasars, N_{QSO} , is $\sim 2,000$. For number densities in the literature¹²⁻¹⁶, the calculated number of pairs is $\sim 800 f_{\text{rc}}^2$ for rich clusters, $50 f_{\text{cg}}^2$ for compact groups, and $5 f_{\text{gg}}^2$ for groups in general. The above numbers are uncertain because the quasars have not been observed in a homogeneous way (n_{QSO} depends on limiting absolute magnitude and redshift, taken here to be $M_B = -25.5$ and $z \approx 1.5$). For specificity, we assume that the number density of galaxy associations is constant in co-moving coordinates, and the quasars evolve as given by their observed densities at these redshifts¹⁶.

We conclude that the discovery of three physical pairs of quasars in galaxy groups is plausible based on the known densities of quasars and galaxy groups. For example, if $f_{\text{rc}} \approx 0.05$ (like galaxies) and $f_{\text{groups}} \geq 0.3$ (as for nearby quasars), then several quasar pairs may be expected in rich clusters and in groups.

What do we expect to observe for quasar groups? (1) Quasar pairs would be more likely than higher multiplicities (such as the triples predicted by the gravitational lens hypothesis) because of the rarity of quasars. (2) Quasar pairs in groups similar to the nearby quasar-galaxy groups⁶⁻⁸ would have projected separations in the range ~ 1 arc s to ~ 1 arc min (for $z \geq 1$). The smaller separations would be favoured if both quasars were massive and near the centre of the group. Separations of ≤ 1 arc s are not expected because typical separations among galaxies in groups are ≥ 10 kpc. By contrast, the gravitational lens hypothesis requires many lensed images with smaller separations. Separations as large as a few arc minutes, as observed for 1146+111, could correspond to quasar pairs in large groups or rich clusters. (3) The velocity difference between the two quasars (probably among the most massive members of the system) may be less than or equal to the typical velocity difference between galaxies in a group, which is a few hundred⁶ kilometres per second ($\sim 100 \text{ km s}^{-1}$ for compact groups¹⁴). (4) The continuum and emission-line spectra are not expected to be identical for the members of the quasar pair. Some evidence of mutual photoionization might be discernible. However, the spectra should be at least as similar as for unrelated quasars at the same redshift. It would be useful to obtain high signal-to-noise spectra for a large number of quasars, in order to develop quantitative statistical measures of the similarity of spectra of physically distinct quasars. (5) In contrast to what is expected on the basis of the lens hypothesis, there is no requirement that background objects be lensed, nor that there be correlated time variations (on the lens timescale). (6) The Zeldovich-Sunyaev effect would not be expected to be large enough to be measured with current techniques. (7) An association of galaxies at the same redshift as the quasars might be found on extremely deep exposures of the field of 1146+111.

We have examined the published data for 1635+267, 2345+007 and 1146+111. For these objects, the existing information is insufficient to distinguish between a single object with two images formed by a large unseen mass or a pair of quasars

belonging to an association of galaxies. However, a pair of images with no visible prominent lens is more simply interpreted as a pair of quasars rather than the lensing of a single object. Of the three systems considered here, 1146+111 is the most likely candidate for a physical pair because its separation is much greater than that for any of the confirmed⁵ gravitational lenses.

This work was supported in part by NSF grants PHY-8217352 and NAS 5-29225, and by the Exxon Corporation. We thank J. E. Gunn, J. B. Oke and J. G. Hoessel for providing us with a copy of the CCD frame of 1146+111.

Received 6 June; accepted 15 July 1986.

1. Djorgovski, S. & Spinrad, H. *Astrophys. J.* **282**, L1-4 (1984).
2. Weedman, D., Weymann, R. J., Green, R. F. & Heckman, T. M. *Astrophys. J.* **255**, L5-9 (1982).
3. Turner, E. L. *et al. Nature* **321**, 142-144 (1986).
4. Burke, W. L. *Astrophys. J.* **244**, L1 (1981).
5. Turner, E. L. *IAU Symp.* No. 117, 209-218 (1986).
6. Stockton, A. *Astrophys. J.* **223**, 747-757 (1978).
7. Stockton, A. *IAU Symp.* No. 92, 89-97 (1979).
8. Stockton, A. *Astrophys. J.* **257**, 33-39 (1982).
9. Gunn, J. E. *et al. Bull. Am. astr. Soc.* **16**, 477 (1984).
10. Schneider, D. P., Gunn, J. E. & Hoessel, J. G. *Astrophys. J.* **264**, 337-355 (1986).
11. Schneider, D. P., Dressler, A. & Gunn, J. E. *Astr. J.* (in the press).
12. Bahcall, N. A. *Astrophys. J.* **232**, 689-698 (1979).
13. Bahcall, N. A. *Astrophys. J.* **270**, 20-38 (1983).
14. Williams, B. A. & Rood, H. R. *Astrophys. J.* (in the press) (1986).
15. Hickson, P. *Astrophys. J.* **255**, 382-391 (1982).
16. Schmidt, M. & Green, R. F. *Astrophys. J.* **269**, 352-374 (1983).

Simulating the sunspot cycle

R. N. Bracewell

Space, Telecommunications and Radioscience Laboratory, Stanford University, Stanford, California 94305, USA

Fossil records of presumed solar activity^{1,2} in 680-Myr-old varves reveal two periods, ~ 314 and 350 varve years respectively, which also appear in recent sunspot records. These periodicities, combined with a nonlinear feature of sunspot number and a theory of the asymmetry of the solar cycle, enable the sunspot record to be extrapolated back to 1800 so well that the next magnetic cycle may be predicted. Fossil records offer the hope of going beyond numerical analysis into physics; one result is the demonstration that part of the 8-15 yr spread of the 11-yr cycle is due to shifts in the year of sunspot minimum caused by the 350-yr cycle. The annual sunspot number can be represented by³:

$$R_{\text{sim}}(t) = |\mathcal{N}[\text{Re}\{E(t) \exp[i(\omega_0 t + \phi(t))]\} + U(t)]| \quad (1)$$

where ω_0 is the angular frequency corresponding to the magnetic period (~ 22 yr), $E(t)$ is an instantaneous envelope amplitude, $\phi(t)$ is the associated instantaneous phase, $U(t)$ is an additive undulation of low amplitude, $\mathcal{N}$ is a nonlinear function, and $R_{\text{sim}}(t)$ is the estimated annual mean sunspot number. Here I consider these parameters, in turn, to derive an evaluation of equation (1) from AD 1800 to 2000.

The magnetic cycle, generally described as having a 22-yr period, consists of two '11-yr' semicycles. Equation (1) indicates that the durations of successive semicycles should diverge oppositely from the mean. Thus in odd-numbered semicycles, such as cycle 19 which began in 1954.2, the negative bias constituted by $U(t)$ advances the start and retards the finish, making the semicycle long, while the same bias has affected cycle 20 in the opposite sense. (For this interpretation, one adopts the notion of an alternating oscillation⁴, whose topology harmonizes with an internal 22-yr torsional oscillator⁵.) The duration of the magnetic cycle should thus be more stable⁶ than that of the semicycles, an implication confirmed by Fig. 1. Part of the spread in semicycle length is therefore due to a magnetic field that adds algebraically, equally in both hemispheres, to the sunspot-

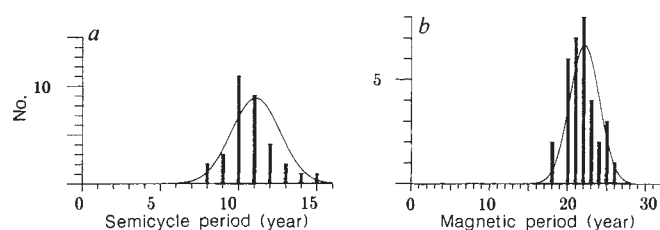


Fig. 1 Duration histograms for sunspot semicycles (a) and magnetic cycles (b). Means and standard deviations are 11.1 ± 1.15 and 22.2 ± 2.08 , respectively.

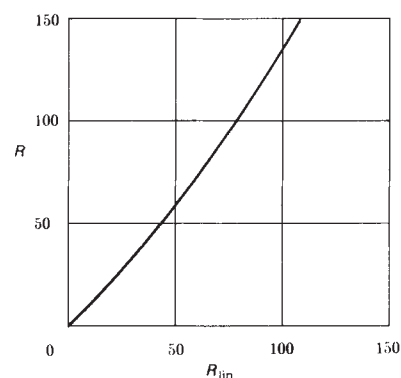


Fig. 2 Nonlinear dependence of sunspot number R on linearized value R_{lin} .

generating field. With mean duration $T_M = 2\pi/\omega_0 = 21.85$ yr we get $\omega_0 = 2\pi/21.85 = 0.2876$ rad yr $^{-1}$. Taking 1986.5 as the provisional epoch of the current sunspot minimum, we adopt $M(\tau) = 73 \sin[2\pi(\tau - 1986.5)/T_M]$ for the magnetic oscillation. The factor 73 anticipates the nonlinear amplitude dependence to be treated below; the time variable τ will also be discussed.

The instantaneous envelope for the recorded sunspot numbers since 1700 has been determined by Hilbert transform analysis³. If used for simulation, neither this envelope nor Dicke's⁷ would contribute to prediction because the fit to historical data is guaranteed but contains no assurance about the future. The envelope^{6,8} for fossil solar cycles, referred to as the Elatina cycle, had a mean period of 314 yr. The non-sinusoidal waveform $E_E(t)$ had a main peak lasting ~ 70 yr to half level, preceded by a smaller peak ~ 100 yr before. Fourier analysis gives the empirical representation $E_E(t) = 1 + \sum (a_n \cos n\theta + b_n \sin n\theta)$, where $\theta = 2\pi t/T_E$, $T_E = 314$ yr, and the coefficients for $n = 1$ to 7 are $a_n = 0.343, 0.044, 0.073, 0.062, 0.017, 0.000, 0.003$; $b_n = 0.095, 0.185, 0.085, 0.037, 0.024, 0.013, 0.007$. Is the Elatina cycle still active? Taking the strong solar activity of recent decades to be associated^{9,10} with the main peak of the Elatina cycle, we use $E(t) = E_E(t - 1956)$, getting the year 1956 from the instantaneous envelope of the sunspot number.

The instantaneous phase $\phi(t)$ of the time-varying analytic function¹¹ describing the complex envelope was allowed for as follows. Changes in magnetic cycle frequency are accompanied by a change of strength in the direction that would be expected³ if the frequency change were a Doppler effect. This kind of Doppler effect is familiar in acoustics where a fixed siren is heard at a fixed point through a medium that is in motion. It has also been found^{9,10} that the 'thickest varves tend to coincide with relatively short cycle-periods,' which strengthens the Doppler interpretation, as this connection persisted over 1,337 varve yr. An adjustment to the year of sunspot minimum can be incorporated into the simulated curve using a regression coefficient based on the correlation that was found³. For the 186-yr interval, which is only part of one Elatina cycle, a linear phase drift, which provides a reasonable adjustment, was incorporated by adopting a magnetic period $T_M = 21.85$ yr, and forcing a minimum in 1986.

The additive undulation seen in the varves^{1,8} is seen in the sunspot record today. The 350-yr period cannot be accurately confirmed, because of the shortness and uneven quality of the modern record, but the magnitude was at a turning point, with a value of -10 , around 1910³. The simulation assumes the mean period T_U of 350 yr that was noted in the Precambrian; thus, $U(t) = -10 \cos 2\pi[(t - 1910)/T_U]$, where $T_U = 350$ yr.

In the Precambrian, the average length of the magnetic cycle was 24 varve yr. Therefore, it might be surmised that what was once 314 is now 288, and what was 350 is now 304. That would be the case if 24 had become 22 solely because of an 8% increase in the Earth's period of revolution over the course of 6.8×10^8 yr (which is $\sim 5\%$ of the age of the Universe and 15% of the age of the Sun). A cosmological decrease in the universal gravita-

tional constant G , as was originally mooted by Dirac^{12,13}, would work in the required direction¹⁴, but tight limits have been placed by radar measurements of the orbit of Mercury¹⁵. Moreover, if G has changed, the internal constitution of the Sun has changed, and it is unlikely that the long periods would have been unaffected except as measured in Earth revolutions. Other influences have been at work on the Earth's period of revolution, as attested by the synchronism between the rotation period of Venus and the synodic period of revolution of Venus as seen from the moving Earth¹⁶, and the Sun has evolved. As the mechanisms of the two periods are unknown, it would be premature to adjust T_E and T_U , and the present results do not in themselves provide a basis for adjustment. However, modern values for T_E and T_U are potentially extricable from polar ice cores and sea sediments. Between 1505 and 1710, an interval which bridges the Elatina peaks, the period seen in both ¹³Be concentration in the Milcent core and thermoluminescence in a Tyrrhenian Sea core is 10.75 yr. Between 1880 and 1915, which embraces a low-amplitude stretch of the Elatina cycle and the minor peak, a period of 11.4 yr is found¹⁷. These values accord with the Doppler shift expected from the 314-yr cycle. Therefore such studies could determine the modern status of both the Elatina cycle (T_E) and the undulation (T_U).

A nonlinear amplitude effect³ exists in the sunspot record which, if removed, makes the oscillations of sunspot number more nearly sinusoidal. This important nonlinearity, which sharpens the semicycle waveform, the more so as the amplitude is greater, revealed itself through a third harmonic (period = 22/3 yr), which was never detected by conventional spectral or correlogram analysis of the sunspot series in its customary (rectified) form, but was reported by Cole¹⁸, the first author to Fourier-analyse the alternating (de-rectified) series. The spectrum of rectified varve-thickness measurements^{9,10} does not show this spectral feature because of cancellation by the method of analysis. The law of nonlinearity can be quantified roughly by $R = \mathcal{H}(R_{lin}) = R_{lin} + \beta R_{lin}^2$, where $\beta = 0.0035$ (Fig. 2). This converts a sunspot number $R = 135$ to a linearized value $R_{lin} = 100$ —a substantial change, but small sunspot numbers suffer only a small correction. The factor 73 adopted for the magnetic cycle $M(t)$ corresponds to a sunspot number 92, which represents an average maximum sunspot number.

Waldmeier¹⁹ established that strong semicycles have short rise times relative to the semicycle duration. The phenomenon resembles the steepening of an acoustic waveform, where the fast rise and slow fall result from the higher speed of a pressure crest. If the sunspot cycle is viewed as the arrival at the photosphere of an upward-travelling wave of which azimuthal magnetic field is a component, then Waldmeier's relation would follow if a crest of magnetic intensity travelled faster, as might result from the magnetic buoyancy^{20,21} associated with density reduction in the presence of magnetic pressure (which varies with the square of the magnetic field). Magnetic buoyancy has

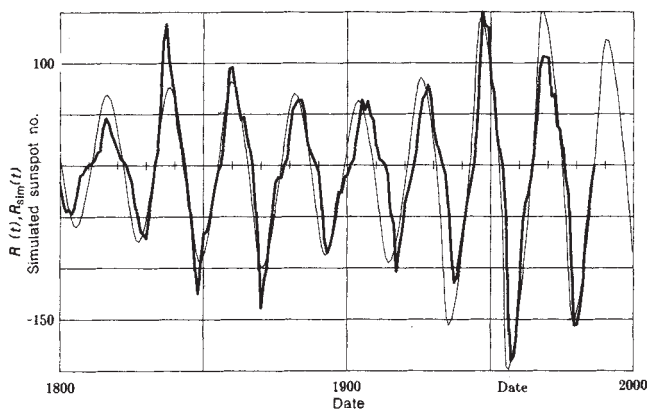


Fig. 3 Observed sunspot number $R(t)$ (heavy curve) and the simulating function $R_{\text{sim}}(t)$.

also been suggested²² as being responsible for strong cycles transporting their magnetic flux to the surface sooner; here, on the contrary, I assume that magnetic buoyancy leaves the spacing of sunspot minima unaffected but produces asymmetry. In a quantitative characterization, the arrival time of a crest is advanced in proportion to the square of the sunspot number. Events equally spaced in time at the level of the submerged generator of the solar cycle will appear at the surface as if time is progressing non-uniformly. The relation between generator time τ and surface time t is taken as $t = \tau - \alpha R^2 + t_{\text{trans}}$, where R is the sunspot number and t_{trans} is a constant delay due to transit. The coefficient α cannot be calculated from a theory of the assumed travelling wave, but its magnitude is determined from the records of cycle shapes, given the non-uniform time equation, and will constrain future theory. The time relation adopted is $t = \tau - 1.5(R/150)^2$ (yr). The magnetic cycle $M(\tau)$, which is deemed to be monochromatic at its source, will be expressed in surface time t as $M(t + \alpha R^2)$. Because of their long periods, $E_E(t)$ and $U(t)$ are not appreciably affected by the distinction between t and τ .

Combining the magnetic cycle $M(\tau)$ with the additive undulation $U(t)$ and multiplying by the Elatina envelope factor $E_E(t)$, we form the function $R_{\text{lin}}(t) = E_E(t - 1956)[M(t + \alpha R^2) + U(t)]$. Applying the nonlinear relations and taking the absolute value, we arrive at the simulated sunspot number $R_{\text{sim}}(t) = |R_{\text{lin}} + \alpha R_0^2 \text{sgn}(R_{\text{lin}})|$.

Figure 3 compares $R_{\text{sim}}(t)$ with the Zürich sunspot number $R(t)$ for the period 1800 to the present. The quality of fit is extraordinarily good, considering that it has never before been possible to explain this curve, except by empirical fit. The broad detail of the Wolf modulation (sometimes groundlessly³ referred to as an 80-yr cycle) is generally reproduced; this is because the secondary peak of the Elatina cycle occurred around 1850 and had the right width; the main peak, placed intentionally at 1956, also proved to have the right width. Consequently, the fact that the Wolf modulation is reasonably reproduced supports the idea that the Elatina cycle is still running with more or less the same waveform and period.

A second type of success is exhibited by the zig-zag alternation of strong and weak maxima. From 1840 to the present the zig-zag effect is quite closely followed. This is partly due to the intentional fixing of the epoch of the undulation for maximum effect in 1910 (a date that could probably be improved), but it is clear that the undulation still persists.

Features of disagreement are seen in the values of the maxima reached in 1937 and 1968, occasions when the observed number of sunspots failed to reach the simulated level. The opposite is seen for 1837, 1848 and 1870. However, the other 12 maxima were simulated to a precision well within the intrinsic range of

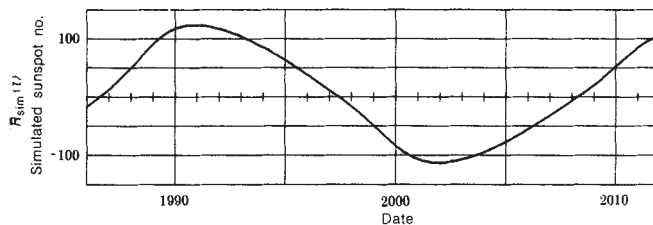


Fig. 4 Future sunspot numbers.

variation, and a scatter diagram of simulated versus observed maxima would show a correlation coefficient of >0.9 . Agreement with the nineteenth century numbers is good and can be interpreted as a vindication of the methods used by R. Wolf in converting diverse sunspot reports to Zürich sunspot number. The simulation fails to reproduce the Maunder minimum^{23,24}. A further discrepancy is disagreement in years of minimum, which partly results from accepting a linear approximation to $\phi(t)$. Some discrepancy may also arise from non-solar influences on the Elatina cycle. Meanwhile, the results obtained, even without refining $\phi(t)$, are good enough to inspire some confidence in projection into the future.

In Fig. 3, part of $R_{\text{sim}}(t)$ extends into the future. Figure 4 shows the coming magnetic cycle on an expanded timescale, until AD 2012. The provisional dates use a minimum epoch of 1986.5 (the precise epoch of the minimum will not be fixed until 1987). Judging from the agreement between 1800 and 1956, the prediction may fit future sunspot counts to within a standard deviation of 20, especially when the starting epoch is adjusted.

Knowledge of the long-term amplitude modulation and additive undulation which are evident from the fossil record (but not independently determinable from the historical sunspot record because of its brevity) has made it possible to reproduce many peculiar features of the sunspot record. An amplitude-modulating envelope of 314-yr period and distinctive waveform, and a 350-yr additive undulation which existed in sediments for thousands of years in Precambrian time, both appear in sunspots today. The results are compatible with parameters that are more or less unchanged in 6.8×10^8 yr.

Two outstanding questions concern the origins of the 350- and 314-yr periodicities. In view of the persistence of these variations a meteorological explanation of the long periods seen in the Precambrian is ruled out, although there may be some degree of meteorological or other non-solar influence. To originate in the Sun, the magnetic field of the 350-yr undulation must be opposite in the north and south hemispheres, and must reverse sign in both hemispheres at intervals of $\frac{1}{2}T_U$. A wave mode³ which can account for this special topology receives support to the extent that the undulation is non-terrestrial. The amplitude modulation has been connected with magnetic cycle length and attributed to the Doppler effect. A second contribution to the spread of semicycle length comes from real advances and delays of sunspot minimum due to the undulation. Neither the two characteristic peaks nor the 314-yr period of the Elatina cycle have been explained, except generally, in terms of the putative convective motion that produces the Doppler effect. The discovery of fossil sequences may now allow the identification of the basic physical mechanisms.

I thank G. E. Williams for his superb analysis, R. G. Giovanelli for introducing me to the rock sample that stimulated the present work, and Philip H. Scherrer for comments on the manuscript.

Received 4 April; accepted 1 August 1986.

1. Williams, G. E. *Nature* **291**, 624–628 (1981).
2. Williams, G. E. *Scient. Am.* **255**, 88–96 (1986).
3. Bracewell, R. N. *Aust. J. Phys.* **38**, 1009–1025 (1985).
4. Bracewell, R. N. *Nature* **171**, 649–650 (1953).

5. Menzel, D. H. *Our Sun* (Harvard University Press, Cambridge, 1959).
6. Bracewell, R. N. & Williams, G. E. *Mon. Not. R. astr. Soc.* (in the press).
7. Dicke, R. H. *Nature* **280**, 24-27 (1979).
8. Williams, G. E. *Aust. J. Phys.* **38**, 1027-1043 (1985).
9. Williams, G. E. in *Weather and Climate Responses to Solar Variations*, 517-533 (ed. McCormack, B. M.) (Colorado Associated University Press, Boulder, 1983).
10. Williams, G. E. & Sonett, C. P. *Nature* **318**, 523-527 (1985).
11. Born, M. & Wolf, E. *Principles of Optics* (Pergamon, New York, 1959).
12. Dirac, P. *Nature* **139**, 323 (1937).
13. Dirac, P. *Proc. R. Soc. A* **165**, 199-208 (1938).
14. Williams, G. E. *Benchmark Papers in Geology*, 57 (Hutchinson Ross, Stroudsburg, 1981).
15. Shapiro, I. I., Smith, W. B., Ash, M. B., Ingalls, R. P. & Pettengill, G. H. *Phys. Rev. Lett.* **26**, 27-30 (1971).
16. Evans, J. V. & Hagfors, T. *Radar Astronomy* (McGraw-Hill, New York, 1968).
17. Castagnoli, G. C., Bonino, G., Attolini, M. R., Galli, M. & Beer, J. *Nuovo Cim.* **7C**, 235-244 (1984).
18. Cole, T. W. *Sol. Phys.* **30**, 103-110 (1973).
19. Waldmeier, M. *Astr. Mitt., Zürich* **14**, 105-130 (1935).
20. Parker, E. N. *Astrophys. J.* **121**, 491-507 (1955).
21. Parker, E. N. *Cosmical Magnetic Fields* (Clarendon, Oxford, 1979).
22. Dicke, R. H. *Nature* **276**, 676-680 (1978).
23. Eddy, J. *Science* **192**, 1189-1202 (1976).
24. Eddy, J. *Scient. Am.* **246**, 80-92 (1977).

The effect of temperature on shuttle glow

G. R. Swenson, S. B. Mende & E. J. Llewellyn*

Lockheed Palo Alto Research Laboratories, 3251 Hanover Street,
Palo Alto, California 94304, USA

The glow on ram surfaces of the space shuttle has been reported from a series of photographic observations made during several orbiter missions¹⁻¹³. These measurements have shown that the spectrum of the glow is a continuum^{7,13}, has a spectral peak at 680 nm^{4,7}, and the brightness decreases with altitude⁵. The spectrum has been tentatively identified as the nitrogen dioxide continuum and follows the interaction of adsorbed nitric oxide with ramming atmospheric oxygen⁷⁻⁹. One recent observation of the ram glow displayed an unusually low intensity^{10,15}. Further investigation has led us to discover that the main difference between this, and earlier measurements was the spacecraft attitude which in turn significantly influenced the temperature of the ram surfaces observed. In this paper the variation of the glow brightness among several different shuttle flights is re-examined and it is shown that a major contributing factor to glow brightness is probably the temperature of the rammed surface. The derived temperature dependence is also consistent with the Atmospheric Explorer-C satellite 'red' glow intensity data.

Laboratory observations of the heterogeneous recombination of oxygen atoms have shown that the molecular yield is a function of temperature¹⁴ and that the surface reaction with different materials allows the preferential formation of particular excited states of O₂¹⁵. The *in-situ* recombination of atomic oxygen to form O₂, presumably on a surface, is a standard technique for the measurement of atomic oxygen concentrations with satellite mass spectrometers. A similar temperature effect in heterogeneous recombination, although in this case for NO₂, has been reported by Engebretson and co-workers for the mass spectrometer on the DE-2 spacecraft¹⁶⁻¹⁸. These investigators were able to operate the satellite mass spectrometer in either a hot, or cold, mode and they noted that the signals due to NO₂ and NO were substantially affected by raising and lowering the temperature of the surfaces of the ion source. It has been postulated that the shuttle glow is a result of oxygen and nitrogen atoms being swept out from the atmosphere, by the spacecraft, forming NO on the surfaces, some of which is adsorbed and undergoes subsequent bombardment by atmospheric O to form NO₂⁷⁻⁹. The formation of NO by surface catalytic processes has

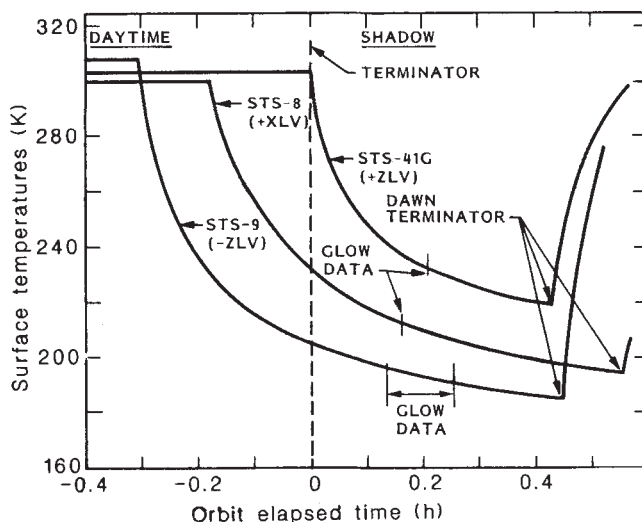


Fig. 1 A plot of surface temperature versus time for glow measurements made on the STS-8, STS-9 and STS-41G missions. The time scale is given with respect to the time the orbit crosses the terminator into orbital darkness. Note that the STS-9 attitude manoeuvre, approximately 0.3 hours before crossing the terminator, resulted in the viewed ram surface being shadowed from the Sun and a very cold surface during glow measurements. The STS-41G attitude maintained sunlight on the ram surface until the terminator crossing. The attitude was bay earthward and resulted in less cooling prior to the glow measurement.

been reported in laboratory experiments¹⁹. The recent mass spectrometer studies^{17,18} have suggested that the physical process postulated for the shuttle glow is a probable explanation for the variations in the mass spectrometer signals corresponding to masses 30 and 46. Thus, it is reasonable to expect that the shuttle glow brightness will be similarly affected by the temperature of the surface exhibiting glow.

The surface of the spacecraft (shuttle) for which glow measurements have been made include the tiles on the vertical stabilizer and OMS pod located on the aft portions of the orbiter. All measurements have been made within the Earth's shadow portion of the orbit (orbital night). The tile surface of the orbiter is an extremely good thermal insulator, for conduction from its mounting substrate, and has an emissivity close to unity. When the orbiter is in an 'airplane' mode, with the stabilizer pointed away from the Earth and not illuminated by sunlight, the studied glow surfaces cool extremely quickly. For an attitude with these same surfaces pointed earthward, the Earth radiates to these surfaces and prevents them from extreme cooling. Solar illumination of a surface prior to entering the terminator will maintain it at a high temperature, but if on the dayside of the orbit the surface is shadowed by the spacecraft itself, and exposed to space, a cold surface temperature can be obtained. Thus the temperature of the glowing surface, at the time of the observations, will depend on both the initial temperature before the spacecraft crosses the terminator, and the subsequent orientation of the spacecraft.

A detailed thermal analysis of the vertical stabilizer has been undertaken for the attitudes and environmental conditions appropriate to the glow measurements on the missions STS-8, STS-9 (Spacelab 1) and STS-41G. The glow observations from the 41G mission are somewhat unusual as they represent a low glow luminosity at low altitude. The results of the thermal analyses are presented in Fig. 1. It should be noted that the general temperature pattern indeed fits that expected. For STS-9, the shuttle is in the 'air-plane' mode so that the stabilizer is radiating to space. As the orbiter had assumed an attitude where the surface to be put in ram was shadowed from the Sun, from

* Permanent address: Institute of Space and Atmospheric Studies, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

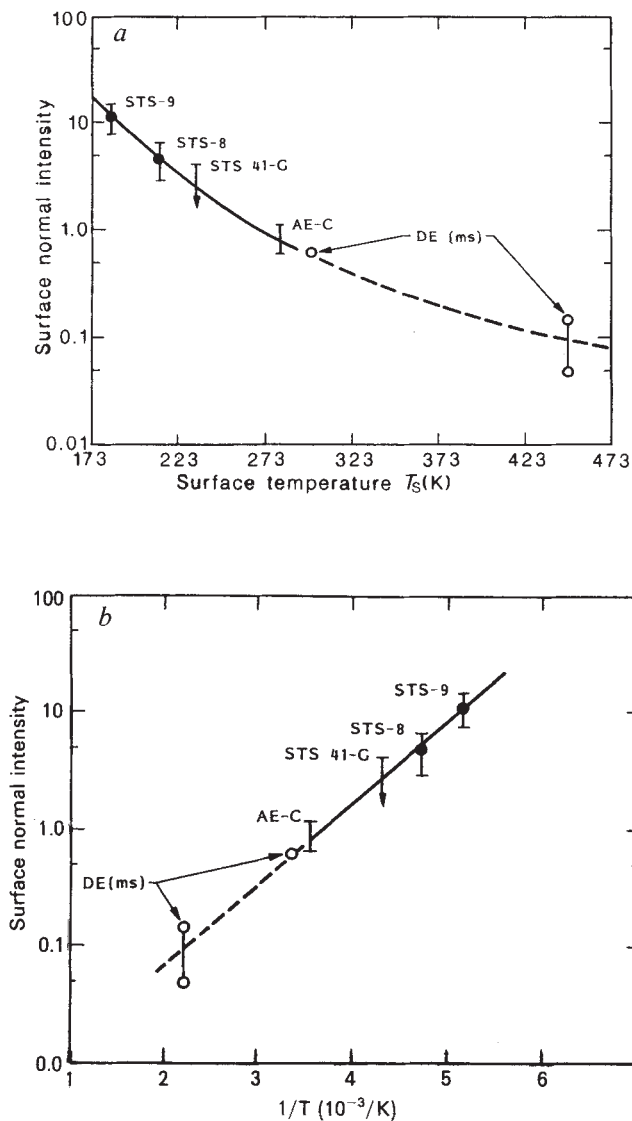


Fig. 2 *a*, A plot of surface normal glow intensity (normalized to 250 km) for the STS-8, STS-9 and STS-41G observations. Units on x axis are R (Rayleighs; 10^6 photons $\text{cm}^{-2} \text{sec}^{-1}$) per $\text{\AA}$ at a wavelength of $7,000 \text{ \AA}$. The solid curve through the data is a best fit curve, exponential, with an equivalent bond energy of 0.14 eV . The dashed portion of the curve is the extension of the same formulation. The DE(ms) points on the plot are the inferred slope of the temperature effect from the Dynamics Explorer (Mass Spectrometer) measurements. The inferred slope, $2.64 \times 10^{-3} [\exp(1.625 T_s)]$, was deduced from the NO_2 signal reported at the cold and warm temperature cycles for the same altitude. *b*, A plot of intensity, (normalized to 250 km), versus $1/T_s$ for the data presented in *a*. The straight line fit was used to deduce the 0.14 eV surface bond energy for NO_2 : $I = I_0 [\exp(E/kT_s)]$ where $E = 0.14 \text{ eV}$.

well before entering the terminator, the glow measurements were made following a long cooling period. For STS-41G, the orbiter was inverted so that the stabilizer received radiation from, and radiated to, the Earth. In addition, the ram surface had been illuminated prior to the point of terminator crossing so that it was initially warm. For the STS-8 mission the attitude is intermediate between these two extremes with the stabilizer radiating towards both space and the Earth.

If, as suggested above, the intensity of the shuttle glow is a function of temperature then the glow intensities measured on

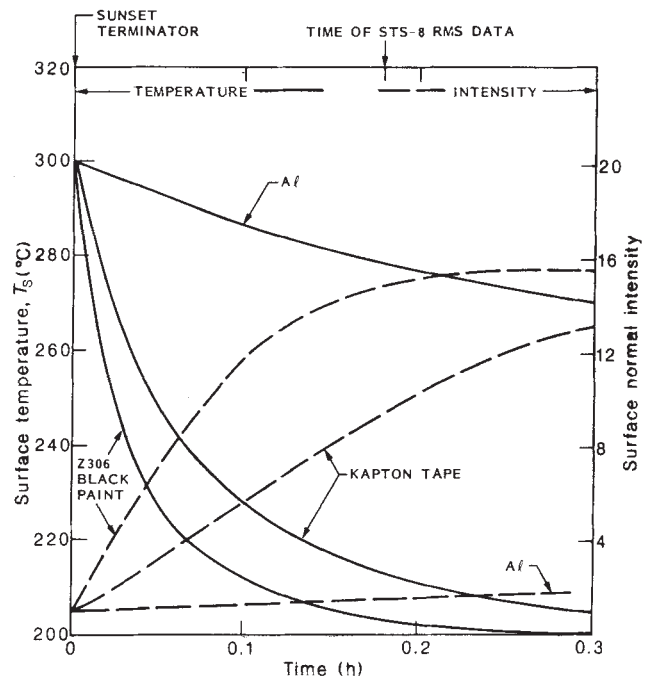


Fig. 3 A plot of temperature versus time (solid curve) for the three foil samples (Kapton, carbon paint and aluminium) flown on the STS-8 mission. The plot of the expected intensities (R per $\text{\AA}$ at 222 km) for each material versus time (dashed curve) was derived with the expression described in Fig. 2*a* for the material surface temperature (T_s) dependence.

these different missions are of interest. However, as the glow brightness, at least for the AE-C satellite, has been shown to be a strong function of altitude⁵, a direct comparison of the different measurements requires that all measurements be reduced to a single common altitude. For the shuttle the precise nature of the variation of the glow with altitude is uncertain, although for the Atmospheric Explorer measurements it has been shown that the glow is a linear function of the ambient atomic oxygen density, at least for altitudes above 180 km ²⁰⁻²². A comparison of the glow observations from the missions STS-3, 4 and 5 qualitatively confirms that the shuttle glow follows a similar altitude variation⁵. The STS-3, 4 and 5 analysis of altitude variation was made from a comparison of uncalibrated images taken with identical instrumentation and with the same shuttle orientation, stabilizer radiating to space. We have, therefore, scaled the measured intensities to a common altitude, 250 km , with the atomic oxygen density derived from the MSIS-83²³ model atmosphere appropriate for the time of the measurement. It should be noted that the altitudes of the glow measurements for STS-8, STS-9 and STS-41G were 222 , 250 and 230 km and the scaling factors to normalize the data to 250 km are 0.57 , 1.0 and 0.67 respectively. These correction factors are much less than the observed variations. Unfortunately the database of calibrated shuttle glow images is quite limited as many observations are contaminated by thruster firings, and we have included all measurements for which there are reliable glow intensities. A correction of the viewing geometry has also been applied to the different measurements as the observations were not made with the same path-length across the vertical stabilizer¹². Finally an additional correction for the ram angle effect was also included¹².

The available glow intensity data from three shuttle missions (STS-8; 9 and 41G) are plotted in Fig. 2*a*, corrected as described above, as a function of glowing surface temperature. For completeness we have also included glow information from the AE-C spacecraft²²; the adopted temperature is that of the photo-

Fig. 4 An example of the material sample glows obtained with the hand-held imager from STS-8 mission. The brightest glow is associated with the carbon paint while the weakest is from the aluminium foil.



meter baffle (J. H. Yee and V. J. Abreu, private communication). It is clear from earlier analysis of the glow e-fold²⁰ and spectrum²¹ for the AE-C satellite that these attributes are different from those observed for the STS spacecraft. The AE-C glow analysis is complicated by the nature of the baffle geometry and protrusion of the baffle beyond the spacecraft skin. These differences remain unreconciled. The solid line in Fig. 2a is a best fit curve to the glow versus temperature data. We believe that there is additional information in the temperature variation of the NO₂ signal recorded by the mass spectrometer on the DE-2 spacecraft. It has been suggested that the physical processes responsible for forming NO₂ in the mass spectrometer, and those producing the NO₂ leading to shuttle glow are the same^{17,18}. The decrease in the NO₂ signal associated with a wall temperature increase in the DE-2 mass spectrometer shows the same trend as the decrease in shuttle glow intensity versus temperature. This is illustrated by the dashed curve in Fig. 2a.

In Fig. 2b the measured glow intensities for the various observations have been plotted against $1/T$. The DE-2 mass spectrometer data were entered on this figure, and on Fig. 2a, with the assumption that the low temperature data were in exact agreement with the best fit curves. The straight line extending through the data is a best straight line and yields a surface bond energy for the NO₂ of 0.14 eV. The expression for the brightness of the glow, resulting from the knowledge of the bond energy, is included in Fig. 2a, for an altitude of 250 km.

In making the foregoing analysis we have necessarily made many assumptions and have neglected the fact that the glow intensity may depend upon the nature of the material surfaces (that is, STS-9, AE-C and DE-2). In practice, the glow from different material surfaces flown on STS-8 and 41D was found to differ according to surface material^{4,8,11}. These material samples had been mounted on a aluminium foil substrate five thousandths of an inch thick which was mounted on a beta cloth-covered, multilayer, insulation blanket. The material samples, like the tile, have a near zero heat source when exposed to deep space at night and cool at a rate which depends on their thermal emissivity.

In Fig. 3 we have plotted the modelled temperatures for samples of Kapton tape, Z306 (polyurethane based carbon paint) and aluminium. These were the three materials flown on the RMS on the STS-8 mission⁴. As the quality of the glow images obtained from this low altitude mission (222 km) is good we believe that the relative brightness of the different glows may be reasonably compared with those expected from our analysis.

Unfortunately, similar measurements from STS-41D were of poor quality due to the high altitude of the mission, 290 km, so these have not been modelled. In Fig. 3, the sample foil temperatures have been plotted as a function of time after terminator crossing. The simple model assumes that the foil temperatures are the same, before entering the terminator, and that the foils are radiating towards both the Earth and deep space. The expression for intensity versus temperature deduced from the shuttle tile measurements shown in Fig. 2a has been used to calculate the intensity expected over each of the foil samples, as a function of time. The relative brightness of the surface samples, as measured with the hand-held imager, is also shown in Fig. 3 and is in good agreement with that modelled. The actual imager photograph is shown in Fig. 4 and on the cover. It is readily apparent from this model that the glow intensity is very dependent on when, in the cooling cycle, the glow measurements are made. For the night-time situation each of the foils is modelled to cool but those with a high emissivity cool most quickly. We, therefore, conclude that it is the emissivity character of the material samples, combined with their thermal capacitance, that produced the variation in glow luminosity over the STS-8 sample plates rather than a specific material surface trait. The cooler surfaces, regardless of material type, would appear to adsorb more NO, which can be converted to NO₂ by ramming atomic oxygen, than warmer ones.

This conclusion does not contradict an earlier hypothesis⁷ that the orbiter glow in the visible region is due to a process involving the surface adsorption of NO and subsequent recombination of ramming atmospheric O on the surface adsorbed NO. The spacecraft must scavenge atmospheric N and O to replace the NO lost to the O beam at a much faster rate than its removal (by atmospheric O). In addition, the agreement of the intensity of the 'red-glow' in the AE data with that predicted from the shuttle observations supports the suggestion that it may also be due to NO₂ (ref. 22).

NO emission above shuttle surfaces has been reviewed in the literature²⁴⁻²⁷. It is of interest to extrapolate these ideas to glow emissions in the infrared and ultraviolet regions of the spectrum, although we do not have shuttle measurements at these wavelengths. If NO is formed efficiently on the surface, then we would expect both gas phase and surface sticking NO as products. For the warmer surfaces the NO would go to gas phase NO, (such as observed in the DE mass spectrometer) with very little sticking to the surface. The gas-phase emission produced by the surface recombination would be expected to

form beta bands¹⁹ (whereas the pure gas reaction results in a mixture of NO beta, gamma, and delta bands)²⁸. It may be that these NO emissions are in fact contributing to the ultraviolet glow signature observed in the 280 and 337 nm channels of Atmospheric Explorer data. Significant infrared emission due to the fundamental and first overtone bands of NO, at 5.3 and 2.7 μm respectively, would also be expected²⁷.

We, therefore, conclude that temperature does play a major role in the surface reactions leading to the shuttle 'ram' glow

and, probably, all spacecraft ram glows.

We thank the many people within NASA who contributed to the experiments and to the post-mission analysis. In particular we extend appreciation to L. Leger, J. Visentine and B. Brown at JSC and R. Gause and colleagues at MSFC for providing much of the necessary thermal and material information used in this study. This analysis has been supported by NASA, Space Telescope and the Lockheed Independent Research Programs.

Received 24 February; accepted 1 August 1986.

1. Banks, P. M., Williamson, P. R. & Raitt, W. J. *Geophys. Res. Lett.* **10**, 118-121 (1983).
2. Mende, S. B., Garriott, O. K. & Banks, P. M. *Geophys. Res. Lett.* **10**, 122-125 (1983).
3. Mende, S. B. *AIAA Rept.* No. 83-CP838, 79-91 (1983).
4. Mende, S. B., Banks, P. M. & Klingelsmith, D. A. *Geophys. Res. Lett.* **11**, 527-530 (1984).
5. Mende, S. B., Nobles, R., Banks, P. M., Garriott, O. K. & Hoffman, J. J. *Spacecraft* **21**, 374-381 (1984).
6. Mende, S. B., Swenson, G. R. & Clifton, K. S. *Science* **225**, 191-193 (1984).
7. Swenson, G. R., Mende, S. B. & Clifton, K. S. *Geophys. Res. Lett.* **12**, 97-100 (1985).
8. Mende, S. B. & Swenson, G. R. *NASA Conf. Publ.* No. 2391, 1-34 (1985).
9. Swenson, G. R., Mende, S. B. & Clifton, K. S. *NASA Conf. Publ.* No. 2391, 35-45 (1985).
10. Kendall, D. J. W. *et al.* *NASA Conf. Publ.* No. 2391, 56-62 (1985).
11. Mende, S. B. *et al.* *J. Spacecraft* **23**, 189-193 (1986).
12. Swenson, G. R., Mende, S. B. & Clifton, K. S. *Geophys. Res. Lett.* **13**, 509-512 (1986).

13. Kendall, D. J. W. *et al.* *Planet. Space Sci.* (submitted).
14. Slanger, T. G. *J. chem. Phys.* **69**, 4779-4791 (1978).
15. Kenner, R. D. & Ogryzlo, E. A. *J. chem. Phys.* **80**, 1-6 (1984).
16. Engebretson, M. J. & Mauersberger, K. *J. geophys. Res.* **84**, 839-844 (1979).
17. Engebretson, M. *NASA Conf. Publ.* No. 2391, 46-54 (1985).
18. Engebretson, M. & Hedin, A. *Geophys. Res. Lett.* **13**, 109-112 (1986).
19. Reeves, R. R., Manella, G. & Harteck, P. *J. chem. Phys.* **32**, 946-947 (1960).
20. Yee, J.-H. & Abreu, V. J. *SPIES Rept.* No. 338, 120-124 (1982).
21. Yee, J.-H. & Abreu, V. J. *Geophys. Res. Lett.* **10**, 126-129 (1983).
22. Yee, J.-H., Abreu, V. J. & Dalgarno, A. *Geophys. Res. Lett.* **12**, 651-654 (1985).
23. Hedin, A. E. *J. geophys. Res.* **88**, 10170-10188 (1983).
24. Green, B. D., Caledonia, G. E. & Rawlins, W. T. *J. Spacecraft* **22**, 500-518 (1985).
25. Green, B. D., Marinelli, W. J. & Rawlins, W. T. *NASA Conf. Publ.* No. 2391, 82-97 (1985).
26. Kofsky, I. L. & Barrett, J. L. *NASA Conf. Publ.* 2391, 155-164 (1985).
27. Barrett, J. L. & Kofsky, I. L. *NASA Conf. Publ.* 2391, 165-168 (1985).
28. Tanaka, Y. *J. chem. Phys.* **22**, 2045-2048 (1954).

Isotopic enrichment by electron exchange

Gerald R. Stevenson, Matthew P. Espe,
Richard C. Reiter & Dean J. Lovett

Department of Chemistry, Illinois State University, Normal,
Illinois 61761, USA

The importance of isotope separation has increased due to the wide use of isotopes in both energy- and life-science-oriented applications¹. This, coupled with our observation² that the solution electron affinities³ of perdeuterated hydrocarbons are less than those of corresponding undeuterated species, led us to investigate the possibility of isotopic enrichment through electron exchange equilibria. Here we report the observation that nitrobenzene containing ¹⁵N has a higher affinity for solvated electrons in liquid ammonia than does nitrobenzene containing the most common isotope of nitrogen (¹⁴N). When an equal-molar mixture of ¹⁴N-nitrobenzene ($\text{PhNO}_2\text{-}^{14}\text{N}$) and ¹⁵N-substituted nitrobenzene ($\text{PhNO}_2\text{-}^{15}\text{N}$) is added to a solution of solvated electrons in liquid ammonia, the anion radical of $\text{PhNO}_2\text{-}^{15}\text{N}$ is formed pre-

dominantly over that of $\text{PhNO}_2\text{-}^{14}\text{N}$. Furthermore, as the chemical and physical properties of the resulting anion radicals ($\text{PhNO}_2\text{-}^{15}\text{N}^{\cdot-}$ and $\text{PhNO}_2\text{-}^{14}\text{N}^{\cdot-}$) are very different from those of the neutral molecules, the electron exchange equilibrium between these two isotopic species can be utilized to enrich samples selectively with respect to the two isotopic species. Analogous results have been observed for the case of benzophenone-¹²C and benzophenone-¹³C (carbonyl ¹³C).

When a solution containing 0.296 mmol $\text{PhNO}_2\text{-}^{15}\text{N}$ and 0.0813 mmol $\text{PhNO}_2\text{-}^{14}\text{N}$ in very dry liquid ammonia is combined with ~ 0.01 mmol of potassium metal, a paramagnetic solution results which contains both anion radicals ($\text{PhNO}_2\text{-}^{14}\text{N}^{\cdot-}$ and $\text{PhNO}_2\text{-}^{15}\text{N}^{\cdot-}$) and both neutral molecules. However, the electron spin resonance (ESR) spectrum of the mixture at -70°C shows that the two isotopically different anion radicals are not present in the same concentration ratio as that of the neutral molecules (Fig. 1). Indeed, computer simulation of the ESR spectrum shows that the ratio of $[\text{PhNO}_2\text{-}^{15}\text{N}^{\cdot-}]/[\text{PhNO}_2\text{-}^{14}\text{N}^{\cdot-}]$ is 7.1:1. This, averaged with two other independent experimental results, shows that the equilibrium constant for the reaction

$$\text{PhNO}_2\text{-}^{14}\text{N}^{\cdot-} + \text{PhNO}_2\text{-}^{15}\text{N} \rightleftharpoons \text{PhNO}_2\text{-}^{14}\text{N} + \text{PhNO}_2\text{-}^{15}\text{N}^{\cdot-} \quad (1)$$

is 2.1 ± 0.1 .

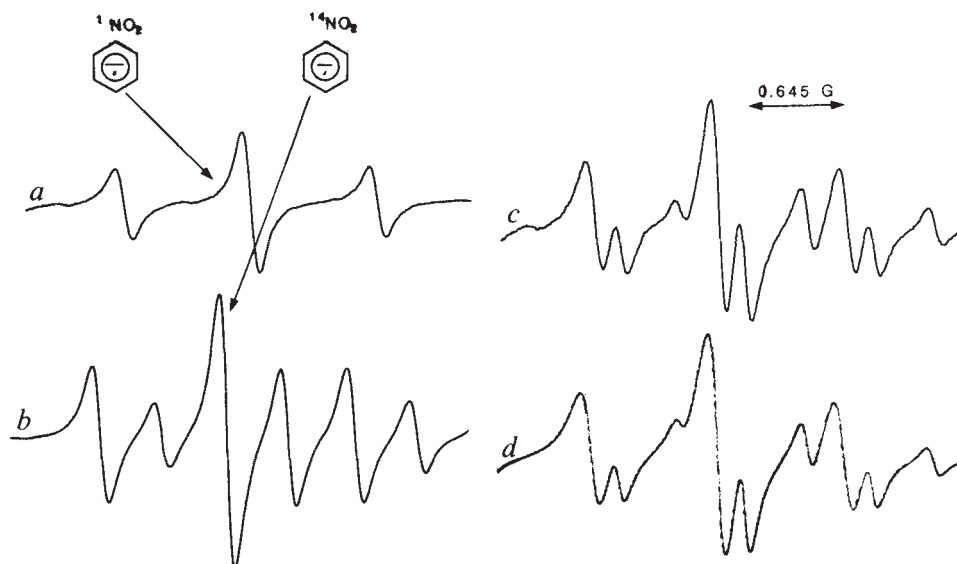


Fig. 1 First three ESR lines of $\text{PhNO}_2\text{-}^{15}\text{N}^{\cdot-}$ (a) and first six ESR lines of $\text{PhNO}_2\text{-}^{14}\text{N}^{\cdot-}$ (b) positioned immediately to the left of the spectral centre. Spectrum c was taken from a 3.5:1.0 mixture of $[\text{PhNO}_2\text{-}^{15}\text{N}]/[\text{PhNO}_2\text{-}^{14}\text{N}]$. The best computer simulation (d) uses a ratio of $[\text{PhNO}_2\text{-}^{15}\text{N}^{\cdot-}]/[\text{PhNO}_2\text{-}^{14}\text{N}^{\cdot-}]$ of 7.1:1.

Evaporation of the ammonia from the solution described above leaves the solid potassium salts of the two anion radicals and the neutral nitrobenzenes in the reaction apparatus. The neutral nitrobenzenes can be easily distilled from this mixture under high vacuum, leaving only the solid $\text{K}^+\text{PhNO}_2^{-14}\text{N}^-$ and $\text{K}^+\text{PhNO}_2^{-15}\text{N}^-$. Recovery of the isotopically mixed nitrobenzenes is easily accomplished by reacting the solid anion radical salts with iodine dissolved in diethylether. The extra electrons are transferred from the anion radical to the I_2 to form I^- as shown in reaction (2):



Because the mixture of solid anion radicals was enriched in $\text{PhNO}_2^{-15}\text{N}^-$ relative to the original mixture of isotopes, the neutral nitrobenzenes remaining after the I_2 addition are similarly enriched in $\text{PhNO}_2^{-15}\text{N}$. Thus, the addition of a deficient number of electrons to a mixture of isotopically substituted nitrobenzenes, followed by separation of the anions from the remaining neutral species, leaves a ^{15}N -enriched nitrobenzene anion radical. The ^{15}N -enriched neutral species can be recovered by adding iodine to this mixture of anion radical salts. The enriched mixture can be further enriched in ^{15}N by simply subjecting this new mixture repeatedly to the same process. Starting from ordinary nitrobenzene containing 0.37% ^{15}N (natural abundance), it would take 16 passes through this procedure (assuming the efficiency is constant) to produce a sample of 99% pure $\text{PhNO}_2^{-15}\text{N}$.

The diminished electron affinity of $\text{PhNO}_2^{-14}\text{N}$ relative to that of $\text{PhNO}_2^{-15}\text{N}$ comes from two effects: an electronic effect and a zero-point energy effect. It has been observed that the substitution of a hydrogen by a deuterium in benzene removes the degeneracy of the symmetric and antisymmetric wave functions⁴. Furthermore, alkyl substituents on the benzene anion radical have the same, albeit stronger, effect, and the alkyl substitution leads to a diminished solution electron affinity⁵. Delvin⁶ has noted that the presence of an added electron in tetracyanoethylene can alter the zero-point energies of the C-C and C-N stretching modes by as much as 25 cm^{-1} . As the nitrogen atom in nitrobenzene is bound to three other atoms and the spin density on the nitrogen atom in PhNO_2^- is large (the nitrogen coupling constant is 10.2 Gauss), this zero-point energy effect could account for a significant portion of the enhanced electron affinity caused by the replacement of ^{14}N by ^{15}N in nitrobenzene.

We believe that the behaviour we have observed in the isotopically different nitrobenzenes is common for systems where the electron density is large on the isotopically substituted atom. This is supported by the fact that our reductions of mixtures of benzophenone- ^{12}C and benzophenone- ^{13}C (BZO- ^{12}C and BZO- ^{13}C) in hexamethylphosphoramide with potassium metal,



show that the solution electron affinity of the BZO- ^{12}C is significantly greater than that of BZO- ^{13}C ($K_{\text{eq}} = 0.58 \pm 0.04$ at 10°C).

The observations reported for both ^{15}N and ^{13}C systems represent an isotope effect on solution electron affinity which is an order of magnitude larger than the usual isotope effects, and is not significantly dependent on the mass ratio, as is usually the case. This suggests the feasibility of isotopic enrichment processes based on partial reduction of mixtures of neutral isotopic species followed by simple physical or chemical separation of the resulting ions from the neutral species in the resulting equilibrium mixtures. Such processes, having unprecedented separation factors coupled with the advantages of working in solution, could make enrichment of a large variety of isotopic species much more practical than is the case today.

We thank the NSF for support.

Note added in proof: the presence of sodium ions (even from the glass) can result in a decrease in the equilibrium constant for reaction (1) to well below unity.

Received 25 March; accepted 14 July 1986.

1. Tanaka, N., Yamaguchi, A. & Arki, M. *J. Am. chem. Soc.* **107**, 7781-7782 (1985).
2. Stevenson, G. R., Espe, M. P. & Reiter, R. C. *J. Am. chem. Soc.* **108**, 532-533 (1986).
3. Stevenson, G. R. & Hashim, R. T. *J. Am. chem. Soc.* **107**, 5794-5795 (1985).
4. Lawler, R. D. & Fraenkel, G. K. *J. chem. Phys.* **49**, 1126-1139 (1969).
5. Lawler, R. G. & Tabit, C. T. *J. Am. chem. Soc.* **91**, 5671-5672 (1969).
6. Khatkale, M. S. & Devlin, J. P. *J. chem. Phys.* **70**, 1851-1859 (1979).

Evaporative cooling of the western equatorial Pacific Ocean by anomalous winds

G. Meyers*, J. R. Donguy† & R. K. Reed‡

*CSIRO Division of Oceanography, Castray Espanade, Hobart, Tasmania 7000, Australia

†ORSTOM Groupe SURTROPAC, BP A5, Noumea, New Caledonia

‡NOAA Pacific Marine Environmental Laboratory, Seattle, Washington 98115, USA

Global climate anomalies during El Niño/Southern Oscillation (ENSO) episodes are controlled by anomalous patterns of sea surface temperature (SST) in the equatorial Pacific Ocean¹⁻³. Many studies during the past decade have indicated that warming of the eastern Pacific is caused by advection and downwelling associated with anomalous eastward currents¹⁻⁶. Cooling of the western Pacific is probably not caused by an analogous process because zonal temperature gradients are small west of the dateline. We have searched for an explanation of the cooling because relatively small temperature changes in the west can be important in influencing the atmospheric general circulation^{3,7,8}. Here we compare oceanic heat storage observed by expendable bathythermographs during 1979-83 with local processes in the heat budget, including various surface fluxes and mixing. The results show that cooling during 1982-83 was caused by evaporation due to anomalous meridional wind. The anomalous wind field in the region had been noted earlier by Harrison⁹.

Ocean temperatures on meridional transects across the western Pacific between New Caledonia (or New Zealand) and Japan have been sampled since 1979 by volunteer observers on merchant ships, who take temperature soundings by XBT (expendable bathythermograph) to 450 m at 100-km intervals along the ship's track. The general transect is typically repeated once or twice a month. Details of the network have been given previously¹⁰.

The surface temperature is coolest at the ends of the transect, where annual oscillation is the dominant variability (Fig. 1). The seasonal range is $\sim 5-7^\circ\text{C}$ at higher latitudes. These oscillations are virtually extinguished within 10° of the Equator, where temperature usually exceeds 29°C . During the first 3 years of our observations, this warm water mass was a stable feature. Then, beginning in June 1982, the warmest waters cooled by $>1^\circ\text{C}$, while the regular seasonal cooling at higher latitudes penetrated more deeply than normal into the tropics. This cooling episode persisted for 16 months, until the later part of 1983, when the 29°C pool was re-established.

The monthly depth of the 26°C isotherm (D_{26}) and the temperature (T_{26}) averaged vertically between the surface and D_{26} were used to characterize the heat content of the mixed layer in the 600-km band spanning the Equator from 2°S to 4°N . There were about 10 XBT stations in the area each month. Surface cooling during the 1982-83 ENSO episode is seen in T_{26} , with the most rapid cooling during the winter months of each hemisphere: July-August 1982 in the southern winter, then January-February 1983 in the northern winter (Fig. 2a). Shoaling of the thermocline by $\sim 50\text{ m}$ during the 1982-83 ENSO episode as seen in D_{26} (Fig. 2b) will be discussed below.

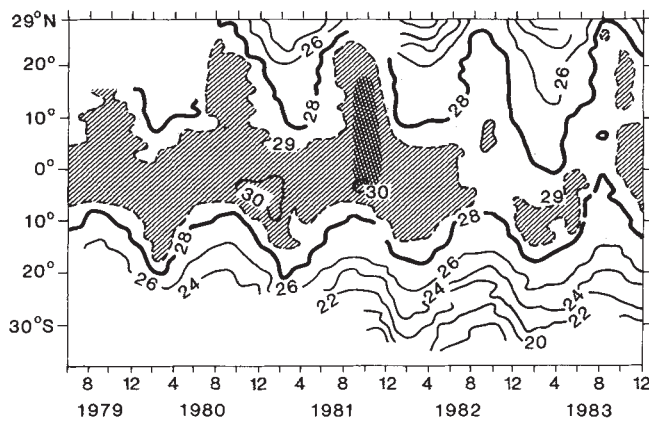


Fig. 1 Sea surface temperature on the shipping route between New Caledonia (or New Zealand) and Japan. Temperatures exceeding 29°C are highlighted with single hatching. Double hatching indicates questionable values due to sparse data.

The processes that influence mixed-layer temperature are surface fluxes, horizontal advection by currents, entrainment mixing, turbulent diffusion, and penetrative radiation at D_{26} (refs 11, 12). The last two of these were found to be small in the equatorial Indian Ocean¹¹, and zonal advection is negligible. In the first model to be tested, the total surface flux (Q_T) is related to the change in upper-layer temperature by the relation

$$\rho C D_{26} \frac{\partial T_{26}}{\partial t} = Q_T \quad (1)$$

where ρ and C are the density and heat capacity of water, and t is time. It is assumed in equation (1) that the mean surface flux is independently balanced by diffusion and advection, thus the equation applies only to fluctuations.

The local rate of heat storage, the left-hand side of equation (1), was calculated from T_{26} and D_{26} , and is typically of the order $\pm 40 \text{ W m}^{-2}$ (Fig. 2f). Periods of persistent cooling seen in T_{26} (Fig. 2a) are associated with the largest negative values of heat storage, as indicated by horizontal brackets. We seek an explanation for these cooling events.

The surface heat fluxes (insolation and net long-wave radiation, latent heat and sensible heat) on the right-hand side of equation (1) can be estimated from empirical formulae, using marine weather observations that include wind speed, SST, wet- and dry-bulb air temperature, and cloud cover. The formulae used in this study have recently been evaluated and are believed to be considerably more reliable than several earlier versions¹³.

The individual marine weather observations, from ship-of-opportunity reports, were first converted to heat fluxes and then averaged monthly for the six 2°-latitude by 10°-longitude areas near the XBT observations (2°S–4°N, 150°–170°E; Fig. 2c–e). The radiative fluxes (insolation minus net long-wave radiation; $Q_S - Q_B$) have a dominant semi-annual timescale recognized previously in other air/sea variables¹⁴. This signal is mainly due to the Sun's passing over the Equator twice a year. The turbulent fluxes (latent plus sensible heat; $Q_E + Q_H$) have a broader range of timescales, with persistent periods of high or low flux tending to last anywhere from a month to almost a year. The total flux Q_T , which is the difference between radiative and turbulent fluxes, is obviously dependent on the timescales and magnitudes of both ($Q_S - Q_B$) and ($Q_E + Q_H$).

A visual comparison of heat storage rate observed by XBT with total surface flux estimated from weather observations, suggests a correlation, particularly during 1982–83 (Fig. 2e, f). The cooling events of largest magnitude, indicated by horizontal bars, occurred during periods when the surface flux was below

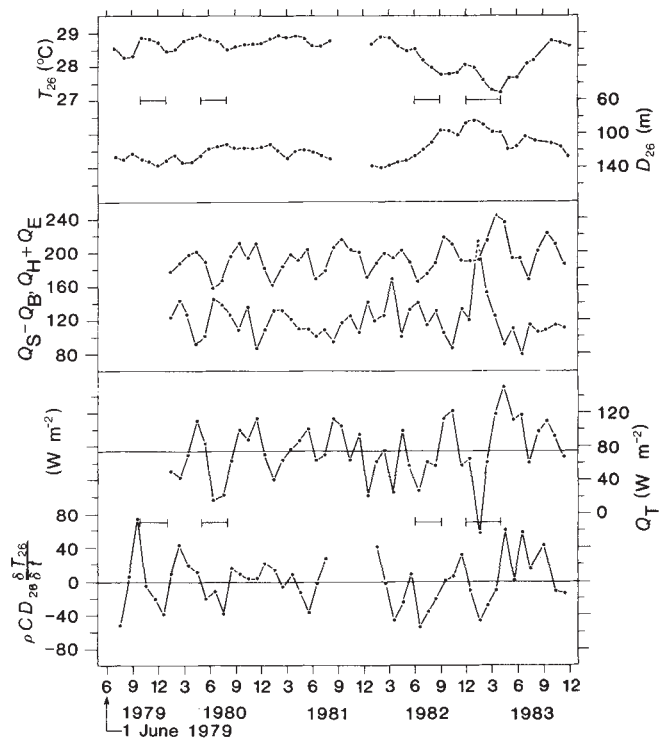


Fig. 2 a, Upper-layer temperature (T_{26}) averaged between the surface and the depth of the 26°C isotherm (b, D_{26}). Heat fluxes at the sea surface: c, net radiation ($Q_S - Q_B$); d, net latent and sensible heat flux ($Q_E + Q_H$); e, total heat flux (Q_T), obtained by subtracting curve d from curve c; f, local heat storage rate ($\rho C D_{26} \partial T_{26} / \partial t$). Horizontal bars bracket times of largest cooling events.

the mean value (75 W m^{-2}) for the 1979–83 period. The correlation coefficient is 0.45, which is significant at the 5% level, allowing for a decorrelation timescale of 2 months.

The next question is, which fluxes have the greatest effect on heat storage rates? Visual inspection of Fig. 2 suggests that changes are due more to irregular fluctuations in turbulent flux than to semi-annual oscillations in radiative flux. The correlations of the heat storage rate [$\rho C D_{26} (\partial T_{26} / \partial t)$] with each of the three heat flux terms (Q_T , $Q_E + Q_H$, and $Q_S - Q_B$) are 0.45, -0.46 and 0.23, respectively. More than 90% of the variance of $Q_E + Q_H$ is due to Q_E . Thus, latent flux is statistically most important, especially during 1982–83.

The weather variables used to calculate $Q_E + Q_H$ are wind speed (W), vertical gradient of vapour pressure ($e_s - e_a$), and vertical gradient of temperature ($T_s - T_a$), where the subscripts s and a indicate values at the sea surface and at 10 m height. The correlation coefficients of $Q_E + Q_H$ with W , $e_s - e_a$ and $T_s - T_a$ are 0.71, 0.26 and 0.47 respectively; they show wind as the dominant factor.

The coefficients given are correlations between turbulent flux and each atmospheric parameter individually. To take into account the relationships between atmospheric parameters, we used principal component analysis¹⁵. The three atmospheric parameters were first normalized by their standard deviations to make them dimensionless, and then the three normalized series were reduced to principal components. This produces three new time series, each one orthogonal to the other, and each a simple linear combination of W , $e_s - e_a$ and $T_s - T_a$. There is no *a priori* reason for any one principal component to be highly correlated to turbulent heat flux; however, one (Q_2) is.

In terms of the original data, Q_2 is calculated from the algebraic equation

$$Q_2 = 0.81 W' + 0.58 (T_s - T_a)' + 0.03 (e_s - e_a)' \quad (2)$$

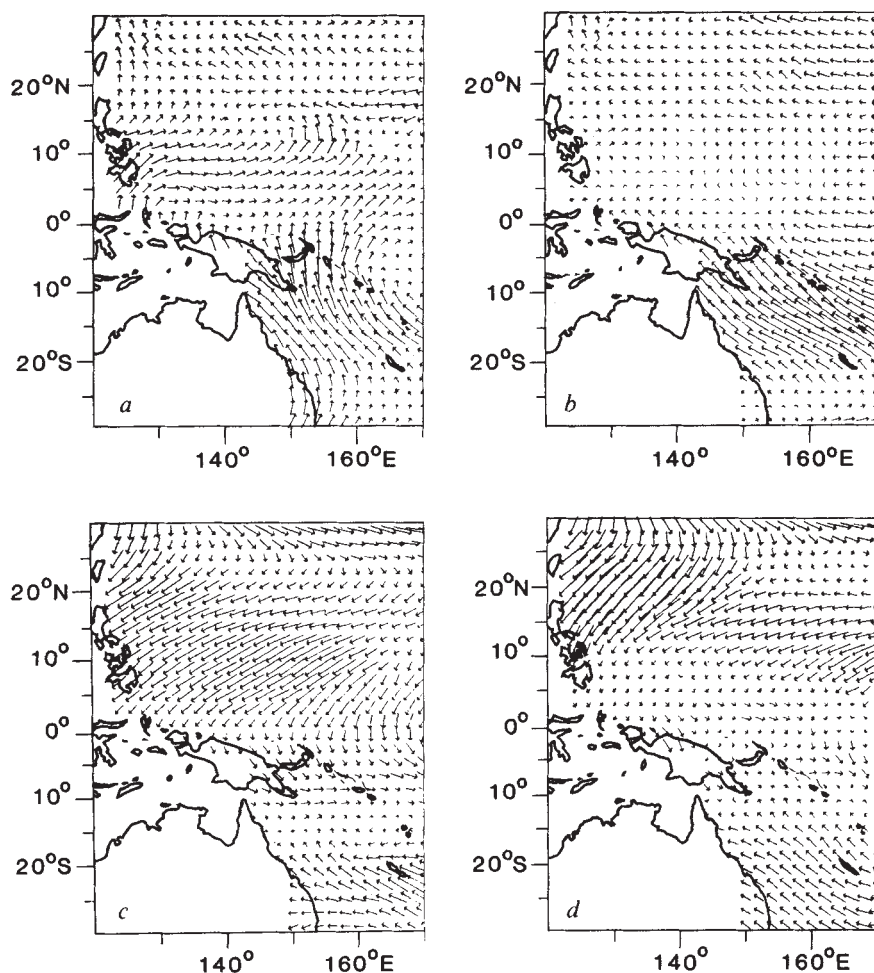


Fig. 3 The field of wind stress at the sea surface during two anomalous cooling periods in 1982–83 (*a*, July 1982; *c*, January 1983) and during two normal periods in 1981 (*b*, July 1981; *d*, January 1981).

where primes indicate that each parameter is measured relative to its mean value and normalized by its standard deviation. Remarkably, the correlation between $Q_E + Q_H$ and Q_2 is 0.89, which accounts for 79% of the turbulent flux variance. This is caused by the structure of Q_2 , in which high wind speeds tend to occur at the same time as large $T_s - T_a$; such a situation increases turbulent flux. The structure of the first principal component (Q_1) had high wind speeds occurring at times of small $T_s - T_a$, producing opposite effects. Neither Q_1 nor the third principal component (Q_3) had a significant correlation with turbulent flux. The weight on W' in equation (2) again indicates that wind speed is the most important governing factor.

The weak dependence of Q_2 and $Q_E + Q_H$ on humidity ($e_s - e_a$) requires an explanation. Generally, humidity is not highly variable over the tropical ocean, as indicated statistically by its coefficient of variation (standard deviation divided by mean value), which is ~ 0.2 . The coefficients of variation for W and $T_s - T_a$ are larger, of the order of unity.

The geographical structure of the cooling wind anomalies can be seen in atlases of Pacific trade-wind fields during recent years^{16–18}. The wind field during July 1982 shows continuous meridional wind blowing from the Tasman Sea across the Equator (Fig. 3*a*). This wind does not penetrate so far equatorward during a normal year, such as 1981 (Fig. 3*b*), when winds near the Equator were weak and variable. The existence of the unusual wind field in 1982 and its effect on ocean currents have been examined by Harrison⁹. The meridional wind also affects the local surface heat budget, as discussed here. Another major cooling event in January 1983 (Fig. 3*c*) again shows a meridional wind blowing across the Equator, this time from the Northern Hemisphere. During a normal year, such as 1981, a weak and

variable wind field is seen near the Equator (Fig. 3*d*). Cooling meridional wind coming from the winter hemisphere is probably a significant factor. Three additional cooling periods before June 1982 (Fig. 2*a*) are associated with strong wind from the winter hemisphere, as seen in the atlases^{16–18}. These are September–December 1979, April–July 1980 and March–April 1982. Thus evaporative cooling at the Equator by anomalous meridional wind probably occurs on several timescales, ranging from brief, random weather inputs¹⁹ to major El Niño episodes²⁰.

Entrainment is a cooling process in which cold water from the pycnocline is mixed into the surface layer. According to McPhaden¹¹, entrainment occurs only when the mixed layer deepens relative to an isotherm near the top of the thermocline. A test of entrainment in the western Pacific seemed necessary because D_{26} shoaled 50 m during the cooling episodes of 1982–83. Estimates of entrainment using McPhaden's method¹¹ for March–September 1982 and December 1982–April 1983 were, respectively, 4 W m^{-2} and 13 W m^{-2} . Thus, cooling by entrainment during the early stage of the episode (4 W m^{-2}) was negligible. Later, after the thermocline had shoaled by 50 m or more, its value (13 W m^{-2}) was still smaller than fluctuations in surface fluxes ($20\text{--}40 \text{ W m}^{-2}$).

Evaporation by the wind has been shown here to be a probable cause of cool SST anomalies in the western equatorial Pacific. This observation identifies surface temperature change controlled by air/sea exchange of heat as part of ENSO. Earlier studies^{1,2,4–6,21} have shown how oceanic heat is redistributed along the Equator, in an adiabatic process that does not involve air/sea heat exchange. This study suggests that the combined action of evaporative cooling west of the dateline and advective warming further east established the anomalous pattern of SST.

John Horel suggested a study of the role of western Pacific temperature in the Southern Oscillation. This research was supported by NSF grants during the Pacific Equatorial Ocean Dynamics (PEQUOD) programme, and by the Equatorial Pacific Ocean Climate Studies (EPOCS) programme of NOAA, and is contribution number 824 from the Pacific Marine Environmental Laboratory.

Received 6 May; accepted 17 July 1986.

- Gill, A. E. & Rasmusson, E. M. *Nature* **306**, 229-234 (1983).
- Cane, M. G. *Science* **222**, 1189-1194 (1983).
- Nicholls, N. *Nature* **307**, 576-577 (1983).
- Wyrtki, K. *J. geophys. Res.* **90**, 7129-7132 (1985).
- Busalacchi, A. J., Takeuchi, K. & O'Brien, J. J. *J. geophys. Res.* **88**, 7551-7562 (1983).
- Harrison, D. E. & Schopf, P. S. *Mon. Weath. Rev.* **112**, 923-933 (1984).
- Palmer, T. N. & Mansfield, D. A. *Nature* **310**, 483-485 (1985).
- Simmons, A. J. Q. *J. R. met. Soc.* **108**, 503-534 (1982).
- Harrison, D. E. *Science* **224**, 1099-1102 (1984).
- Meyers, G. & Donguy, J. R. *Nature* **312**, 258-260 (1984).
- McPhaden, M. J. *J. mar. Res.* **40**, 403-419 (1982).
- Stevenson, J. W. & Niiler, P. P. *J. phys. Oceanogr.* **13**, 1894-1907 (1983).
- Reed, R. K. *J. geophys. Res.* **88**, 9627-9638 (1983).
- Meyers, G. *J. phys. Oceanogr.* **12**, 1161-1168 (1982).
- Lorenz, E. N. *Empirical Orthogonal Functions and Statistical Weather Prediction* (Science Rep. No. 1, Statistical Forecasting Project, MIT Department of Meteorology, Cambridge, 1956).
- Legler, D. M. & O'Brien, J. J. *Atlas of Tropical Pacific Wind-Stress Climatology 1971-80* (Department of Meteorology, Florida State University, Tallahassee, 1984).
- Leetma, A. & Witte, J. (eds) *El Niño Atlas 1982-83* (Nova University Oceanographic Center Dania, FL 33004, 1984).
- Arkin, P. A., Kopman, J. D. & Reynolds, R. W. *1982-1983 El Niño/Southern Oscillation Event Quick Look Atlas* (NOAA National Weather Service, Washington DC, 1983).
- Hasselmann, K. *Tellus* **28**, 473-485 (1976).
- Nicholls, H. *Mon. Weath. Rev.* **112**, 424-432 (1984).
- Philander, S. G. H. & Seigel, A. D. *Proc. 16th int. Liege Colloq. on Ocean Hydrodynamics* (ed. Nihoul, J.) (Elsevier, Amsterdam, 1985).

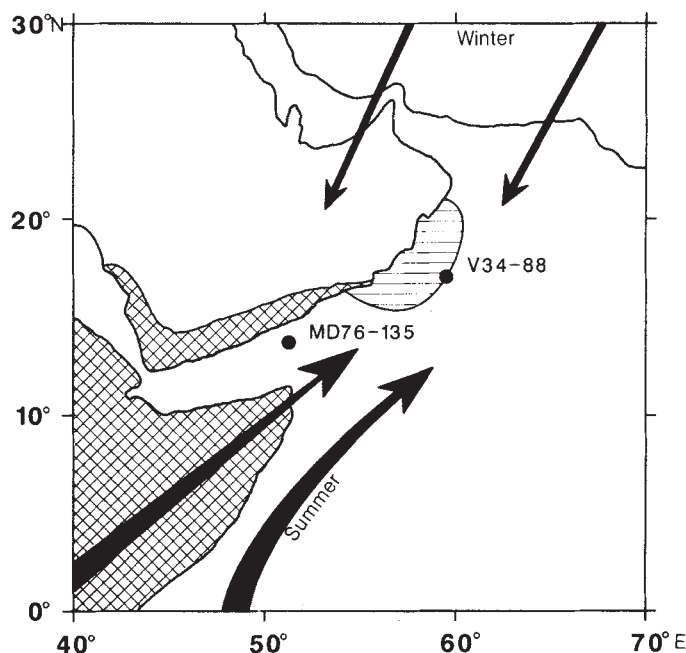


Fig. 1 Location of cores MD76-135 (containing the monsoon pollen index, MPI) and V34-88 (containing the monsoon upwelling index, MUI). The generalized source area for the MPI^{26,27} is indicated by the cross-hatched pattern, and the area of maximum upwelling and MUI^{11,12,23-25} is shown by horizontal shading. Arrows indicate the direction and relative strength of summer southwesterly and winter northeasterly winds.

Coherent response of Arabian Sea upwelling and pollen transport to late Quaternary monsoonal winds

Warren L. Prell* & Elise Van Campo†

* Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA

† Laboratoire de Palynologie, Université de Sciences et Techniques du Languedoc, 34060 Montpellier Cedex, France

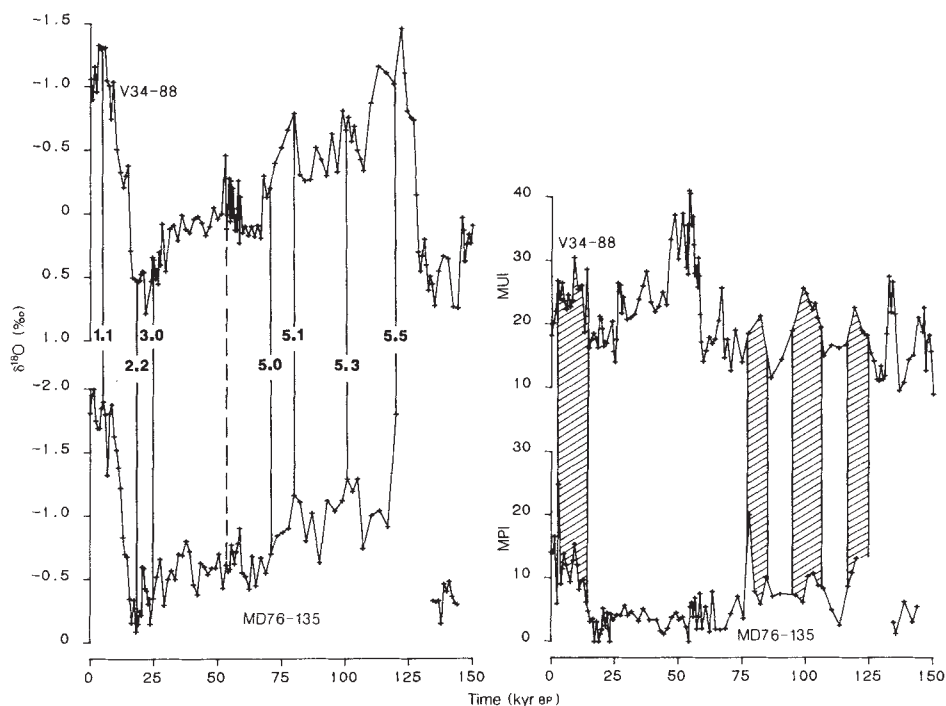
The Indian summer monsoon causes large seasonal changes in the environments of both eastern Africa and the Arabian Sea. Here we compare time series (140 kyr long) of selected pollen types and foraminiferal upwelling faunas preserved in marine sediments, to identify the frequencies of their variability and the coherence and phase of the terrestrial and marine responses to monsoonal circulation. During interglacial intervals, the pollen and upwelling records of the western Arabian Sea indicate stronger monsoons and are both coherent and in phase at periodicities near the precession of the Earth's rotational axis (23 kyr). We conclude that the strength of the monsoonal southwesterlies accounts for this high coherence of marine and terrestrial records during interglacial intervals. Low correlation during glacial phases indicates weaker monsoons and some decoupling of these two records. These data indicate that two distinct modes of monsoon variability may exist: an interglacial mode with strong monsoons and effective coupling of terrestrial and marine records, and a glacial mode with both weaker monsoons and decreased linkage between Arabian Sea pollen and upwelling records.

The atmospheric circulation associated with the Indian summer monsoon reflects both the large-scale differential heating between Africa-Asia and the Indian Ocean, and the complex dynamics of the atmosphere-ocean system^{1,2}. In the western Indian Ocean the summer monsoon circulation is characterized by cross-equatorial transport, strong low-level southwesterly

winds (Fig. 1), coastal upwelling (Fig. 1), and increased precipitation¹⁻⁵. Changes in these climatic variables have been inferred from a large variety of geological data, including palaeolake levels^{6,7}, pollen spectra^{8,9}, wind-transported materials¹⁰, and distribution of planktonic faunas¹¹⁻¹³. These studies indicate that the monsoon was weaker ~18,000 yr ago, during maximum glacial conditions¹⁴⁻¹⁷, and stronger ~9,000 yr ago, when the Northern Hemisphere received ~7% more summer insolation than at present^{4-7,11,17,18}. Although the monsoonal circulation affects both continents and oceans, relatively few studies have compared the responses of these two environments. However, some studies have been able to document the abundance of terrestrial components, such as pollen^{8,9,17}, fresh-water diatoms¹⁰ and phytoliths¹⁰ in marine sediments, and thereby take advantage of the continuous and well-dated nature of marine records. In the western Arabian Sea, records of monsoonal upwelling faunas and wind-transported pollen are preserved in the deep-sea sediments. Our objective here is to determine the coherence of these marine and terrestrial data and to establish their relative response to long-term changes in seasonal solar radiation and glacial boundary conditions (that is, changes of the volume and extent of ice sheets, surface albedo and sea surface temperatures).

We used graphic correlation of oxygen isotopic events¹⁹⁻²¹ to correlate two Arabian Sea cores that contain indices of monsoonal intensity (Fig. 1). Core MD76-135 (14°27' N, 50°31' E; 1,895 m water depth) has an isotopic stratigraphy measured by J. C. Duplessy¹⁶; that of core V34-88 (16° N, 60° E; 2,800 m) was measured by Prell²². The correlation of isotopic stratigraphy between the two cores is good, considering their different accumulation rates (V34-88, 4 cm kyr⁻¹; MD76-135, 8-12 cm kyr⁻¹), the different species used for the isotopic measurements (Fig. 2), and their different depositional environments. The isotopic events were used with the SPECMAP age

Fig. 2 Time series of oxygen isotope stratigraphy and monsoon indicators in Arabian Sea cores. $\delta^{18}\text{O}$ was measured on *Globigerina sacculifer* in core V34-88²² and on *Globigerinoides ruber* in core MD76-135¹⁶. The major isotopic events used for correlation between cores are indicated by the isotopic taxonomy devised by the SPECMAP project¹⁹⁻²¹ and the time scale is from ref. 21. The MUI in V34-88 and the MPI in MD76-135 are scaled in proportion to the faunal and floral populations. Covariance (diagonal hatching) occurs during interglacial rather than glacial intervals.



model²¹ to convert the records to time series. Original sample intervals were ~ 1 kyr in both cores, and the records extend to ~ 144 kyr BP, with a 14-kyr gap in MD76-135 (E.V.C. and W.L.P., in preparation). Figure 2 gives time series of both the $\delta^{18}\text{O}$ stratigraphy and the monsoon indices.

The time series of the monsoonal upwelling index (MUI), which is the per cent *Globigerina bulloides* in the total planktonic foraminifera²², shows maxima centred near 10, 37, 50, 80, 100, 120, 133 and 146 kyr (Fig. 2). High values of the MUI have been associated with coastal upwelling in the western Arabian Sea, on the basis of its spatial coincidence with low sea surface temperature and high nutrient levels^{11,12}. It has been demonstrated²³⁻²⁵ that the southwesterly monsoon winds produce both coastal upwelling and a wide zone of divergence off Arabia which occurs only during the summer months. Thus, stronger summer winds are associated with lower sea surface temperatures, thinner mixed layers and maxima of the MUI. Over the past 150 kyr the MUI has varied with a period of ~ 23 kyr (ref. 22), and is coherent with $\delta^{18}\text{O}$ and orbital variations, although with a slight lag in the latter case.

The time series of the monsoon pollen index (MPI) (from refs 9, 17 and E.V.C. and W.L.P., in preparation) has maxima centred near 10, 80, 100 and 120 kyr (Fig. 2). The MPI is the sum of pollen types that have been collected by shipboard air filters in the western Arabian Sea exclusively during the summer season (ref. 9 and E.V.C. and W.L.P., in preparation), plus other types that can be unambiguously shown to originate from the humid or montane tropical parts of north-east Africa and south-west Arabia^{26,27}. Taxa included in the MPI are *Avicennia*, *Combretaceae*, *Commiphora*, *Indigofera*, *Kohautia*, *Mallotus*, *Olea*, *Phyllanthus*, *Podocarpus*, *Rhizophoraceae*, *Sapotaceae-Meliaceae*, and spores. Maxima of the MPI are interpreted as reflecting both increased monsoonal winds and increased precipitation in the source area, which create more favourable environments for the humid and montane tropical taxa. Although strong monsoonal winds are needed to transport the pollen to the Arabian Sea, the partitioning of the transport and source-area contributions to MPI is difficult to evaluate.

Comparison of the rather different MUI and MPI time series (Fig. 2) reveals that they have synchronous maxima, indicating strong monsoons (shaded areas) during interglacial intervals. However, the MUI contains more maxima and its highest values

occur in isotopic stage 3 (near 50 kyr), where the MPI has no maximum. Overall, the records show large responses that are highly correlated during interglacial stages (stages 1 and 5), but variable responses that are poorly correlated during glacial or intermediate stages (2-4). These data indicate both changes in monsoon intensity and different sensitivity of the monsoon indices. The greater complexity of the MUI record may reflect greater sensitivity of the upwelling system to weaker monsoons relative to pollen transport to the Arabian Sea.

Although the overlapping record length is only ~ 130 kyr, we calculated cross-spectra between the marine and terrestrial monsoonal indices to quantify their frequencies and the coherence between the records. We acknowledge that longer records are preferable, and we view the spectral results as support, rather than proof, of the relationship between the monsoon indices. As these records are relatively short and we are interested in the high-frequency ($\sim (20 \text{ kyr})^{-1}$) variability, we have removed long-term ($(100 \text{ kyr})^{-1}$) trends by linearly detrending the upwelling index in V34-88 and removing the 100-kyr eccentricity component from the pollen index of MD76-135.

We find concentrations of variance density near periods of 40, 23 and 13.3 kyr for the MPI, and near periods of 66 and 23 kyr for the MUI. Unfortunately, the records are not long enough to interpret critically the lower-frequency (> 25 kyr period) peaks (Fig. 3a). The key result is that both spectra have a concentration in variance density over the frequency band centred near $(23 \text{ kyr})^{-1}$ (Fig. 3a). Furthermore, the records are coherent (at the 80% confidence level) over the band of frequencies from $(21 \text{ kyr})^{-1}$ to $(25 \text{ kyr})^{-1}$ (Fig. 3a), and their phase spectrum (Fig. 3b) shows that the upwelling and pollen records are in phase over the coherent frequency bands. We do not interpret the apparently high coherence near 200 and 12 kyr because these peaks are not coincident with concentrations of variance density in the spectra and are thus biased because the spectrum is not effectively constant over the bandwidth²⁸.

These results show that both marine and terrestrial responses to the monsoon contain significant periodicities which are characteristic of the amplification of the seasonal cycle by the precession of the Earth's rotational axis. Additional experiments by us show that the monsoonal indices are coherent (over the $(23 \text{ kyr})^{-1}$ frequency band) with integrated summer (June-August) solar radiation for the Northern Hemisphere, but with

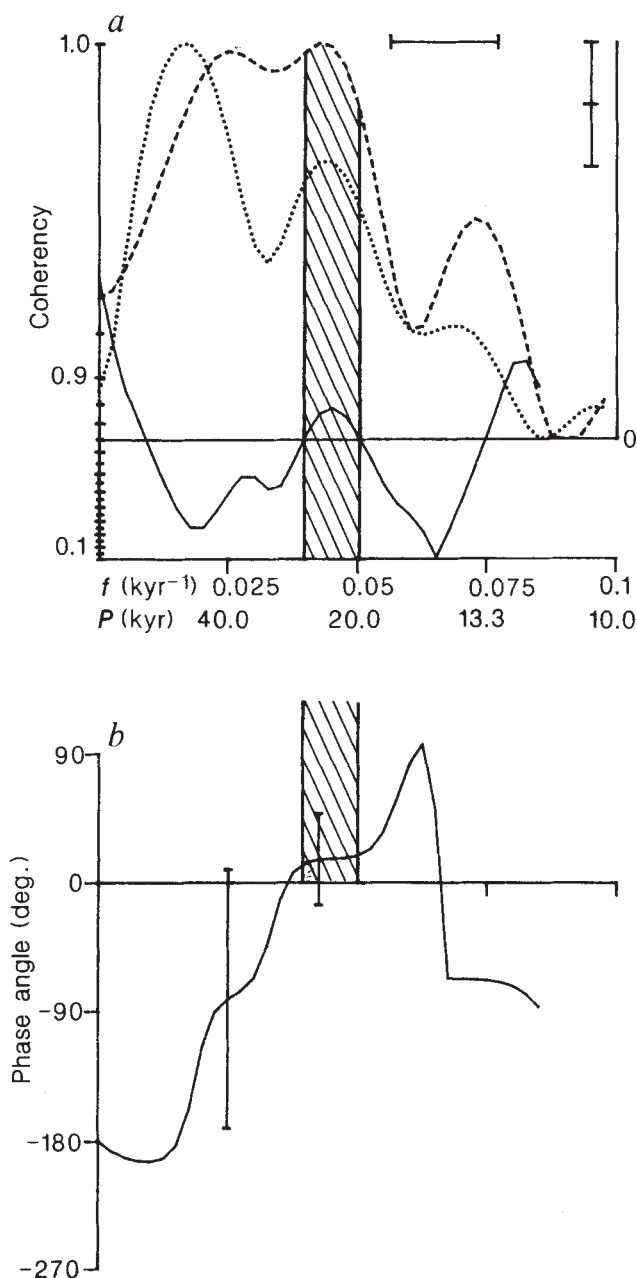


Fig. 3 *a*, Normalized variance spectra of the MPI (dashed line) and MUI (dotted line) and their coherency spectrum (solid line) over the interval from 0 to 144 kyr BP. Variance spectra are plotted on arbitrary logarithmic scales; the coherency spectrum is plotted on a hyperbolic arc tangent scale. The sampling interval is 2 kyr. Horizontal line, critical value of non-zero coherence at the 80% level (0.7). Horizontal bar, bandwidth; vertical bar, 80% confidence interval for variance density. Note that maximum concentrations of variance and coherence occur over the frequency band from $(21 \text{ kyr})^{-1}$ to $(25 \text{ kyr})^{-1}$. *b*, Phase spectrum and estimates of phase error of MPI versus MUI over the interval from 0 to 144 kyr. At $f = (23 \text{ kyr})^{-1}$, 30° in phase corresponds to ~ 2 kyr. Note that MPI and MUI are in phase (phase angle = 0) over the coherent frequency band.

a slight lag (~ 3 kyr). This lag may reflect a slight damping of the monsoon by glacial boundary conditions. However, the time series are not long enough to preclude some non-stationarity of their phase; resolution of the significance of the lag awaits the analysis of longer records.

We interpret these results as indicating that two quite different indicators of the summer monsoon (MPI and MUI) are linked by the same process, the southwesterly winds. This conclusion is modified by the likelihood that higher values of the MPI can partially reflect higher precipitation, resulting in increased vegetation and pollen production in the source areas. The coherency and phase of MPI and MUI suggest that precipitation must be positively correlated (at the 23-kyr periodicity) with stronger monsoonal winds, at least during interglacial phases. This conclusion is supported by general circulation model palaeoclimate simulations^{4,5,7} that incorporate higher solar radiation during the Northern Hemisphere summer, which yield an association between increased summer precipitation in eastern Africa and strengthened southwesterlies in the tropics of the Indian Ocean sector.

During interglacial intervals with strong radiation maxima, both continental and oceanic responses are strong, highly coupled, and reflect the large-scale organization of the regional Indian Ocean monsoonal climate. However, during intervals with weak radiation maxima and substantial glacial boundary conditions, the monsoon indices are different in character and are less coupled. Hence, we suggest that two modes of tropical monsoon variability may occur in response to changes in radiation and glacial boundary conditions. The weak glacial mode reflects weaker monsoons and less effective transport of monsoon pollen to the Arabian Sea; however, upwelling continues during this interval, so the monsoon winds persist but are weaker and may have different trajectories. The strong interglacial mode reflects the direct response of the precipitation and the wind fields to radiation maxima, resulting in increased monsoon-area pollen and effective transport.

We thank J. Kutzbach, T. Webb III and J. Imbrie for helpful discussions, J. C. Duplessy for providing core MO76-135, A. Duffy for running spectral analyses, and the staff at the Center for Climatic Research, University of Wisconsin for their support. This work was supported by Brown University, an NSF grant to E.V.C., NSF grant ATM-8213948 and DOE grant DEFG02-85ER60304 to W.L.P.

Received 6 March; accepted 8 July 1986.

1. Ramage, C. *Monsoon Meteorology* (Academic, New York, 1971).
2. Hastenrath, S. *Climate and Circulation of the Tropics* (Reidel, Dordrecht, 1985).
3. Hahn, D. E. & Manabe, S. *J. Atmos. Sci.* **32**, 1515-1541 (1975).
4. Kutzbach, J. E. *Science* **214**, 59-61 (1981).
5. Kutzbach, J. E. & Guetter, P. J. *J. Atmos. Sci.* (in the press).
6. Street-Perrott, F. A. & Roberts, N. in *Variations in the Global Water Budget* (eds Street-Perrott, F. A., Beran, M. A. & Ratchiffe, R. A. S.) 331-345 (Reidel, Dordrecht, 1983).
7. Kutzbach, J. E. & Street-Perrott, F. A. *Nature* **317**, 130-139 (1985).
8. Rossignol-Strick, M. *Nature* **304**, 46-49 (1983).
9. Van Campo, E. thesis Univ. Sci. Tech. Languedoc (1983).
10. Pokras, E. M. & Mix, A. C. *Quat. Res.* **24**, 137-149 (1985).
11. Prell, W. L. in *Climatic Processes: Sensitivity to Solar Irradiance and CO₂* (eds Hansen, J. E. & Takahashi, T.) 48-57 (American Geophysical Union, Washington, 1984).
12. Cullen, J. L. & Prell, W. L. *M. Micropaleontol.* **9**, 1-52 (1984).
13. Labracherie, M., Barde, M. F., Moyes, J. & Pujos-Lamy, A. in *Sedimentary Records of Ancient Coastal Upwelling*, Pt 13 (eds Thiede, J. & Suess, E.) 347-365 (Plenum, New York, 1983).
14. CLIMAP project members *Geol. Soc. Am. Map Chart Ser. No. Mc-36* (1981).
15. Prell, W. L. *et al. Quat. Res.* **14**, 309-336 (1980).
16. Duplessy, J. C. *Nature* **295**, 49-498 (1982).
17. Van Campo, E., Duplessy, J. C. & Rossignol-Strick, M. *Nature* **296**, 56-59 (1982).
18. Marvil, R. & Prell, W. L. *Geol. Soc. of Am. Abstr. Progr.* **15**, 637 (1983).
19. Prell, W. L. *et al. Paleogeography* **1**, 137-162 (1986).
20. Pisias, N. G. *et al. Mar. Geol.* **56**, 119-136 (1984).
21. Imbrie, J. *et al. in Milankovitch and Climate* (eds Berger, A. & Imbrie, J.) 269-305 (Reidel, Dordrecht, 1984).
22. Prell, W. L. in *Milankovitch and Climate* (eds Berger, A. & Imbrie, J.) 349-366 (Reidel, Dordrecht, 1984).
23. Wyrki, K. *Oceanographic Atlas of the International Indian Ocean Expedition* (NSF, Washington, 1971).
24. Prell, W. L. & Streeter, H. F. *J. mar. Res.* **40**, 143-155 (1982).
25. Luther, M. E. & O'Brien, J. J. *Prog. Oceanogr.* **14**, pp. 353-385 (1985).
26. Zohary, M., *Geobotanical Foundations of the Middle East* (Fisher, Stuttgart, 1983).
27. Keay, R. W. J. *Carte de la Végétation de l'Afrique au Sud du Tropique du Cancer* (Oxford University Press, 1959).
28. Bloomfield, P., *Fourier Analysis of Time Series: an Introduction* (Wiley, New York, 1976).

Humic macromolecule interlayering in nontronite through interaction with phenol monomers

M. C. Wang & P. M. Huang

Department of Soil Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0

Humic substances, which are the principal organic components of soils and waters¹, form complexes with clay minerals²⁻⁴. These complexes profoundly affect the physical, chemical and biological properties of soils and sediments^{4,5}. Various theories on the reaction sites of clay minerals with humic substances have been proposed^{1,5-15}; however, evidence for the retention of humic substances between the layers of expanding layer silicates in field conditions is very limited, and comes only from soils in acidic environments^{10,11,14,16} ($pH < 5$). Studies on the laboratory synthesis of humic substance interlayers in expanding layer silicates have shown that the formation of interlayers was strongly favoured at $pH 2.5$ and diminished significantly between $pH 4$ and 5 (refs 6-9, 12). Aniline polymerizes in interlayers of Fe(III)-saturated montmorillonite but not in those of Ca-saturated montmorillonite¹⁷. Furthermore, aniline is not very representative of biologically generated, natural aromatics. We report here that one of the well identified precursors^{18,19} for the formation of humic substances, hydroquinone, in aqueous solution at near-neutral pH (6.5) can be transformed to humic macromolecules deposited in the interlayers of nontronite (Fe(III)-bearing smectite) saturated with calcium, which is the most common and abundant exchangeable cation of soils and sediments.

The growth of micro-organisms in the systems studied was not evident (Table 1); all the reactions observed in this study were thus abiotic processes. The data (Table 2) show that at the end of a 20-day reaction period, the absorbance of the supernatant of the hydroquinone-nontronite system at 664 nm was 3.0 times higher than that of the hydroquinone system, while at the end of the 11-day reaction period at 472 nm it was 1.8 times higher. The amounts of humic acid (HA) (relative molecular mass $M_r > 1,000$) and fulvic acid (FA) ($M_r > 1,000$) formed in the supernatant of the hydroquinone-nontronite system were, respectively, 1.9 and 3.5 times higher than those in the hydroquinone system (Table 2). This indicates the catalytic effect of nontronite on the oxidative polymerization of hydroquinone and the associated reactions.

After removal of associated organic matter (NaOCl treatment)²⁰ and iron oxides (dithionite-citrate-bicarbonate treatment)²⁰, reoxidation with the NaOCl treatment and saturation

with Ca^{2+} , there is no significant decrease in the catalytic power of nontronite in the polymerization of hydroquinone. Moreover, the sequence of the catalytic role of selected minerals of the smectite group is: nontronite > montmorillonite > hectorite. Nontronite contains Fe(III) in the sheet structure, whereas the major structural cations in the octahedral sheets of montmorillonite and hectorite are Al plus Mg and Mg, respectively. This evidence supports the conclusion that the structural Fe(III) of nontronite plays a vital role in catalysing the formation of humic substances.

The absorption bands of the infrared spectra of the HA ($M_r > 1,000$) and FA ($M_r > 1,000$) were observed at 3,400–3,000 cm^{-1} (hydroxyl groups with varying degrees of H bonding); 2,900–2,800 cm^{-1} (the stretching vibrations of aliphatic CH, CH₂ and CH₃ groups of side chains of aromatic nuclei); 1,715 cm^{-1} (carboxyl as well as aldehydic and ketonic carbonyl); 1,610 cm^{-1} (C=C stretching vibrations in aliphatic and aromatic compounds); 1,380 cm^{-1} (δ_s CH₃, salts of carboxylic acid and/or aliphatic CH); 1,230 cm^{-1} (stretching vibrations of esters, ethers and phenols); and 1,030 cm^{-1} (C–O stretch and ethers). Furthermore, the electron spin resonance (ESR) spectra of the HA and FA formed in the nontronite-hydroquinone system were single lines, devoid of hyperfine splitting. The g -values of HA and FA were 2.0039 and 2.0025, respectively; their breadths between the two extreme peaks on the derivative curves were 5.9 and 5.6. These infrared and ESR spectra resemble those of natural HA and FA extracted from soils (Mollisol and Alfisol in Saskatchewan, Canada). Part of the humic polymers formed was adsorbed on Ca-nontronite to form a humic macromolecule-clay complex with 1.18% of organic carbon content. After the complex was extracted with 0.1 M NaOH, the residue still contained 0.78% organic carbon, although aqueous NaOH is the most widely used solvent for humic substances.

At 58.3% relative humidity and room temperature, Ca-nontronite and the humic macromolecule-Ca-nontronite complex showed the 14.8 Å basal d spacing with almost the same intensity (not shown). The NaOH extraction treatment of the complex did not change the d spacing. On heating at 150 and 200 °C, Ca-nontronite collapsed to yield 11.9 and 10.1 Å basal d spacing, respectively, whereas the X-ray diffractograms of the humic macromolecule-Ca-nontronite complex were very diffuse even after the NaOH treatment (Fig. 1), indicating the deposition of humic macromolecules in interlayers of Ca-nontronite. After heating at 250 and 300 °C, Ca-nontronite had the sharp peak at 9.8 Å, whereas the humic macromolecule-Ca-nontronite complex, even after the NaOH treatment, yielded diffuse weak peaks at 10.0–10.1 Å, which skewed to low angles. The data indicate that the heat treatment at 250 and 300 °C partially decomposed the humic macromolecules adsorbed in the interlayers of Ca nontronite. After heating at 350 °C, the humic macromolecule-

Table 1 Examination of the growth of micro-organisms in hydroquinone-nontronite systems during a 40-day incubation period

Nontronite	Hydroquinone	Thimerosal	Inoculum	Growth of microorganisms in various media		
				GYEN	AIA	Czapek-Dox
+	+	+	–	–	–	–
–	+	+	–	–	–	–
–	–	–	+	+	+	+

One gramme of autoclaved nontronite was suspended in 75 ml of sterilized water in a 150-ml Erlenmeyer flask containing 400 mg hydroquinone and 0.02% (wt/vol.) thimerosal, and then adjusted to $pH 6.5$ with 0.05 M NaOH. The flasks were then plugged with sterilized foam and agitated in a water bath at 25 °C for 20 days. Using a platinum loop, a small portion of each suspension was inoculated to agar plates of GYEN²⁵, AIA²⁶ and Czapek-Dox²⁶ agar at the beginning and end of the reaction period, to examine whether there had been any microbial growth in the systems. All inoculated plates were incubated at 25 °C for 40 days. Thimerosal was used as an antiseptic; it does not affect the oxidation process of phenolic compounds²⁷. The inoculum used was the centrifugate obtained at 1,000g from a soil/water suspension (soil:water = 1:10); the fresh soil was taken from the Ap horizon of an orthic, dark-brown, chernozemic soil (Vertic Agriboroll), a member of the Sutherland Association (106°14' W, 52°18' N). +, In the presence; –, in the absence.

Table 2 Absorbances of the supernatants and formation of humic polymers in hydroquinone–nontronite systems at the end of a 20-day reaction period

	Absorbance		Humic polymers formed in the supernatant (mg)			Humic macromolecules adsorbed on the Ca-nontronite* (mg)	Total humic polymers formed in the reaction system (mg)
	at 472 nm	at 664 nm	HA	FA	Sum		
Hydroquinone plus nontronite	>>2.00 (2.24) [†]	1.12	25	39	64	18	82
Hydroquinone	1.82 (1.28) [†]	0.37	13	11	24	—	24

The flasks which contained the suspensions prepared as described in Table 1 were plugged with sterilized foam and then agitated in a water bath at 25 °C for 20 days. All the experiments were duplicated, including the controls. The pH of the suspension was kept constant by adding NaOH during the reaction period. The suspension in each flask was weighed, and sterilized water was added to keep all suspensions at the same volume. A 5-ml suspension was withdrawn from each flask every, or every other, day and centrifuged at 7,800g for 10 min; the supernatant was then separated and measured for its absorbance at 472 and 664 nm. At the end of the 20-day reaction period, all the suspension in each flask was centrifuged at 7,800g for 10 min to separate the supernatant from the sediment. Supernatants were acidified to pH 1.0 to separate HA and FA, which were then purified^{28–31}. The HA and FA obtained were dialysed against deionized distilled water using dialysis tubes with molecular weight cutoff 1,000. The sediments were extracted with benzene:95% ethanol (1:1) in 1:2 (wt/vol.) ratio²⁷. Part of the residue was taken for organic carbon analysis to estimate the amount of humic macromolecules adsorbed on sediments²⁷. The remaining sediment (humic macromolecule–Ca-nontronite complex) was subjected to X-ray analysis (Fig. 1).

* The organic carbon content (0.31%) of Ca-nontronite was subtracted from that (1.18%) of the humic macromolecule–Ca-nontronite complex; the amount of humic macromolecules adsorbed was then estimated on the basis of the previous finding that synthesized humic substances contain ~50% organic carbon²⁷.

[†] The absorbances of the supernatants at the end of an 11-day reaction period are shown in parentheses.

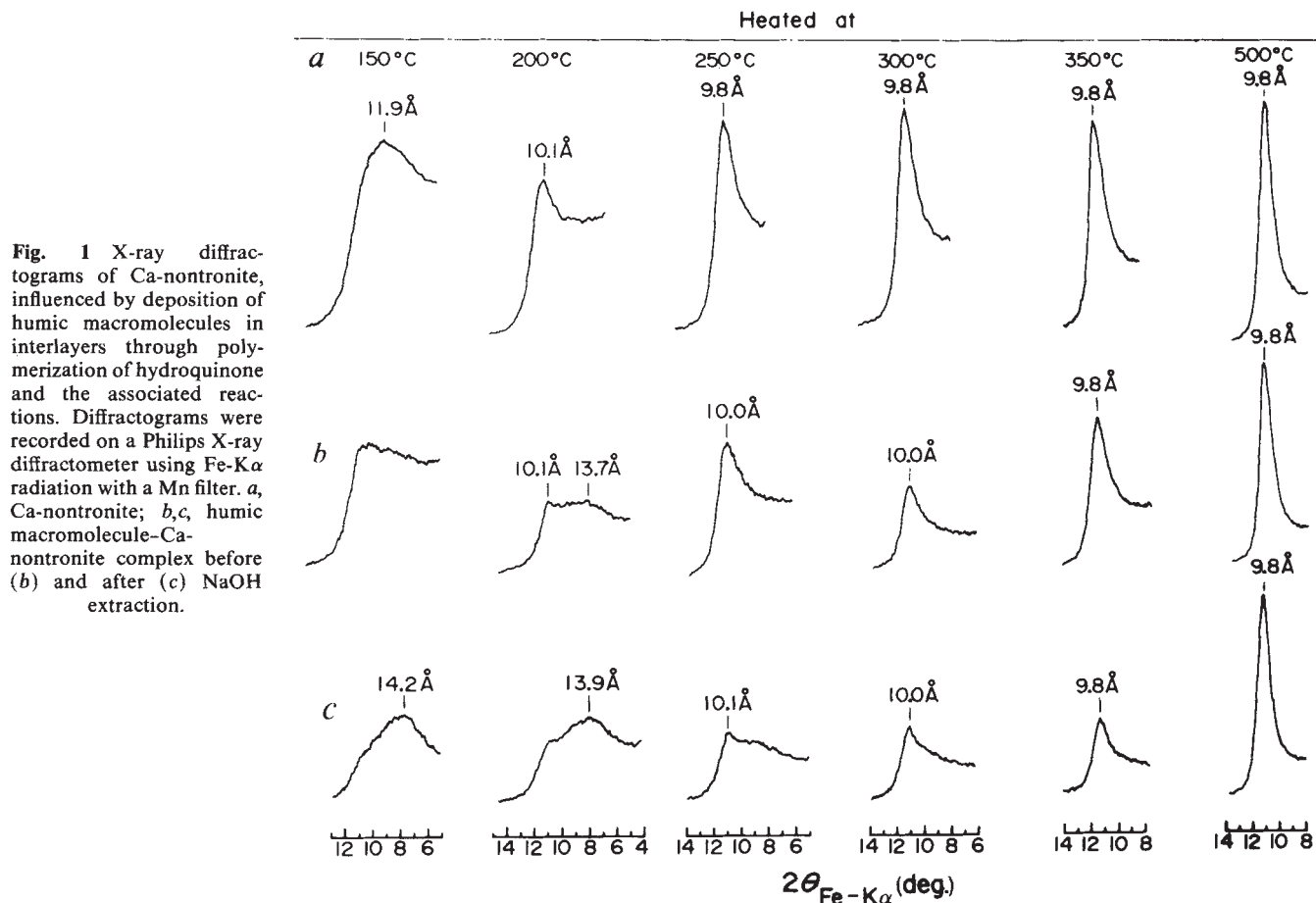


Fig. 1 X-ray diffractograms of Ca-nontronite, influenced by deposition of humic macromolecules in interlayers through polymerization of hydroquinone and the associated reactions. Diffractograms were recorded on a Philips X-ray diffractometer using Fe-K α radiation with a Mn filter. *a*, Ca-nontronite; *b*, *c*, humic macromolecule–Ca-nontronite complex before (*b*) and after (*c*) NaOH extraction.

Ca-nontronite complex before and after the NaOH extraction collapsed to yield 9.8 Å basal *d* spacing, but their X-ray diffractograms were still slightly skewed to low angles, in contrast to the sharp and symmetric X-ray diffraction pattern of Ca-nontronite. This indicates that humic macromolecules present in the interlayers of Ca-nontronite was not completely destroyed even after heating at 350 °C.

The organic carbon contents of the complex after heating at

150, 200, 250, 300, 350, 400 and 500 °C were, respectively, 0.88, 0.46, 0.38, 0.25, 0.13, 0.11 and 0.02% before the NaOH treatment and 0.54, 0.45, 0.29, 0.17, 0.10, 0.08 and 0.01% after the NaOH treatment. The decrease of organic carbon contents of the complex on heating between 150 and 500 °C was accompanied by a decrease in basal *d* spacings. The heat treatment at 500 °C destroyed humic macromolecule interlayers, as indicated by the disappearance of the slight skewness to low angles of the X-ray

pattern (Fig. 1). The changes in the X-ray diffractograms and organic carbon contents of the humic macromolecule-Ca-nontronite complex with increasing heating temperatures clearly show that some humic macromolecules, deposited in the interlayers of Ca-nontronite, can resist heating to $\sim 350^\circ\text{C}$.

Following treatment with benzene:95% ethanol (1:1), the extraction with 0.1 M NaOH removed only 33.9% of the organic carbon from the humic macromolecule-Ca-nontronite complex, indicating that most of the organic carbon adsorbed by nontronite was not solubilized by alkali.

The data show the deposition of humic macromolecules through polymerization and the associated reactions of hydroquinone, a simple and uncharged phenol monomer, in the interlayers of Ca-nontronite at room temperature and near-neutral pH. Most of the interlayer humic macromolecules are highly resistant to the alkali extraction and may thus be humin-type substances. Furthermore, in addition to Al-interlayering of clay²¹⁻²⁴, our present data indicate that the formation of humic substance interlayers in 2:1 expansible phyllosilicates through polymerization of phenol monomers and the associated reactions in soils and sediments merits close attention.

This study was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Received 10 April; accepted 26 June 1986.

- Schnitzer, M. & Khan, S. U. *Humic Substances in the Environment*, 1-7, 253-279 (Dekker, New York, 1972).
- Theng, B. K. G. *Formation and Properties of Clay-Polymer Complexes*, 283-326 (Elsevier, Amsterdam, 1979).
- Schnitzer, M. in *Interactions of Soil Minerals with Natural Organics and Microbes* (eds Huang, P. M. & Schnitzer, M.) (Soil Science Society of America, Madison, in the press).
- Stevenson, F. J. *Humus Chemistry*, 374-402 (Wiley, New York, 1982).
- Schnitzer, M. & Kodama, K. in *Minerals in Soil Environments* (eds Dixon, J. B. & Weed, S. B.) 741-770 (Soil Science Society of America, Madison, 1977).
- Schnitzer, M. & Kodama, H. *Science* **153**, 70-71 (1966).
- Schnitzer, M. & Kodama, H. *Soil Sci. Soc. Am. Proc.* **31**, 632-636 (1967).
- Kodama, H. & Schnitzer, M. *Soil Sci.* **106**, 73-74 (1968).
- Schnitzer, M. *Soil Sci. Soc. Am. Proc.* **33**, 75-81 (1969).
- Kodama, H. & Schnitzer, M. *Can. J. Soil Sci.* **51**, 509-512 (1971).
- Lowe, L. E. & Parasher, C. D. *Can. J. Soil Sci.* **51**, 136-137 (1971).
- Schnitzer, M. & Kodama, H. *Clays Clay Miner.* **20**, 359-367 (1972).
- Kodama, H. & Schnitzer, M. *Trans. 10th int. Congr. Soil Sci. Moscow* **2**, 51-56 (1974).
- Moinereau, J. *Clay Miner.* **12**, 75-82 (1977).
- Theng, B. K. G., Churchman, G. J. in *Abstr. 1985 (8th) int. Clay Conf.*, 238 (Denver, Colorado, 1985).
- Burchill, S., Hayes, M. H. B. & Greenland, D. J. in *The Chemistry of Soil Processes* (eds Greenland, D. J. & Hayes, M. H. B.) 369-400 (Wiley, Chichester, 1981).
- Cloos, P., Badot, C. & Herbillon, A. *Nature* **289**, 391-393 (1981).
- Flaig, W., Beutelspacher, H. & Rietz, E. in *Soil Components Vol. 1* (ed. Gieseking, J. E.) 4-97 (Springer, New York, 1975).
- Hayes, M. H. B. & Swift, R. S. in *The Chemistry of Soil Constituents* (eds Greenland, D. J. & Hayes, M. H. B.) 179-252 (Wiley, Chichester, 1978).
- Jackson, M. L. *Soil Chemical Analysis—Advanced Course* 2nd edn, 37-58 (University of Wisconsin, Madison, 1979).
- Dixon, J. B. & Jackson, M. L. *Science* **129**, 1616-1617 (1959).
- Jackson, M. L. *Clays Clay Miner.* **11**, 29-46 (1962).
- Rich, C. I. *Clays Clay Miner.* **16**, 15-30 (1968).
- Barnhisel, R. I. in *Minerals in Soil Environments* (eds Dixon, J. B. & Weed, S. B.) 331-356 (Soil Science Society of America, Madison, 1977).
- Germida, J. J. & Casida, L. E. Jr *J. Bact.* **144**, 1152-1158 (1980).
- Olson, B. M., McKercher, R. B. & Germida, J. J. *Pl. Soil* **76**, 379-387 (1984).
- Wang, T. S. C., Wang, M. C., Ferng, Y. L. & Huang, P. M. *Soil Sci.* **135**, 350-360 (1983).
- Hayes, M. H. B. in *Humic Substances in Soil, Sediment and Water* (eds Aiken, G. R., McKnight, D. M., Wershaw, R. L. & MacCarthy, P.) 329-362 (Wiley, New York, 1985).
- Aiken, G. R. in *Humic Substances in Soil, Sediment and Water* (eds Aiken, G. R., McKnight, D. M., Wershaw, R. L. & MacCarthy, P.) 363-385 (Wiley, New York, 1985).
- Swift, R. S. in *Humic Substances in Soil, Sediment and Water* (eds Aiken, G. R., McKnight, D. M., Wershaw, R. L. & MacCarthy, P.) 387-408 (Wiley, New York, 1985).
- Leenheer, J. A. in *Humic Substances in Soil, Sediment and Water* (eds Aiken, G. R., McKnight, D. M., Wershaw, R. L. & MacCarthy, P.) 409-429 (Wiley, New York, 1985).

Stable isotope evaluation of the origins of amino acids in fossils

Michael H. Engel* & Stephen A. Macko†

* School of Geology and Geophysics, 830 Van Vleet Oval, The University of Oklahoma, Norman, Oklahoma 73019, USA

† Department of Earth Sciences, Memorial University, St John's, Newfoundland, Canada A1B 3X5

The racemization reactions of amino acids in Quaternary fossil materials have been exploited for absolute dating^{1,2}, stratigraphic correlation^{3,4} and palaeothermometry⁵. Recent work has involved the isolation of amino acid constituents of proteinaceous material from fossil bones for ¹⁴C dating⁶ and for establishing ancient dietary preferences by determining the stable nitrogen and carbon isotopic compositions of fossil collagens^{7,8}. Such investigations are often hindered by a lack of control over the indigeneity of these compounds, as the ubiquitous nature of amino acids generally precludes the determination of the absolute authenticity of amino acids in fossils by conventional means. However, the amino acids in organisms have been found to have distinct stable isotopic compositions^{9,10}. Here we report experimental results which indicate that after the death of an organism, low-temperature diagenetic reactions such as racemization should not cause a significant shift in the stable carbon or nitrogen isotopic compositions of the resultant D- and L-amino acid enantiomer products. We therefore propose that a comparison of these enantiomers from individual amino acids isolated from fossils may provide, for the first time, a method for establishing the absolute indigeneity of these compounds.

Significant developments in analytical techniques, such as gas chromatography (GC), gas chromatography/mass spectrometry (GC/MS) and high-performance liquid chromatography (HPLC), continue to improve the precision with which amino acid D/L values can be determined in fossil materials¹¹⁻¹³. New preparative HPLC procedures have made it possible to isolate

Table 1 Amino acid L/D and D/L values

Time (h)	Starting material: Temperature (°C)	D-Glutamic acid		L-Glutamic acid	
		L/D	K_1 (10^{-4} h^{-1})	D/L	K_1 (10^{-4} h^{-1})
0	100	0.00	—	0.00	—
175	100	0.03	1.71	0.03	1.71
361	100	0.06	1.66	0.06	1.66
554	100	0.09	1.63	0.09	1.63
768	100	0.12	1.57	0.12	1.57
1,392	100	0.21	1.53	0.21	1.53
1,772	100	0.25	1.44	0.25	1.44
96	150	0.67	84.45	0.68	86.57
264	150	0.99	—	1.01	—

The D/L and L/D values have a standard error of ± 0.01 .

individual amino acids from proteins for stable isotopic analyses⁹. The development of accelerator-based mass spectrometric (AMS) methods for the direct determination of ¹⁴C abundances has provided a novel approach for dating small quantities (milligrammes) of protein and amino acids isolated from fossils¹⁴.

Investigations into the effects of peptide sequence¹⁵, environmental parameters (such as temperature, pH, moisture and metal-ion concentration¹⁶) and diagenetic processes (such as hydrolysis, aminolysis, decomposition (decarboxylation, deamination) and polymerization^{6,17}) on amino acid distribution and racemization rates have been beneficial; however, a fundamental problem remains concerning the application of amino acid reactions to geochemical research. Fossils (particularly bone) are, in general, not 'closed' systems. As physical and chemical processes break down a bone matrix, amino acids and peptides are leached out. The ubiquitous nature of amino acids complicates the situation in that, as amino acids are being leached out, non-indigenous amino acids may be introduced by, for example, microbial activity and/or diffusion of water from the surrounding, preserving environment^{6,18}. The prospects for isolating entirely indigenous, proteinaceous material from a fossil for the purposes of dating or stable isotopic analyses can therefore be limited.

Table 2 Stable isotope values of reactants and products

Starting material:		D-Glutamic acid				L-Glutamic acid			
Time (h)	Temperature (°C)	$\delta^{13}\text{C}$ D-Glu (‰)	$\delta^{13}\text{C}$ L-Glu (‰)	$\delta^{15}\text{N}$ D-Glu (‰)	$\delta^{15}\text{N}$ L-Glu (‰)	$\delta^{13}\text{C}$ D-Glu (‰)	$\delta^{13}\text{C}$ L-Glu (‰)	$\delta^{15}\text{N}$ D-Glu (‰)	$\delta^{15}\text{N}$ L-Glu (‰)
0	100	-20.0	—	+7.9	—	—	-25.9	—	-6.2
554	100	-20.2	-20.1	+8.0	ND*	-25.9	-26.2	ND	-5.9
1,392	100	-20.2	-19.8	+7.9	ND	-26.3	-26.1	ND	-5.8
1,772	100	-20.0	-19.7	+7.8	+7.5	-26.0	-25.8	-5.4	-5.9
264	150	-19.8	-20.1	+8.6	+10.0	-26.2	-26.1	-2.1	-4.1

$\delta^{15}\text{N}$ reference = air (0‰); 1 standard deviation = $\pm 0.2\%$. $\delta^{13}\text{C}$ reference = NBS-22 (-29.4%) relative to PDB carbonate (0‰); 1 standard deviation = $\pm 0.1\%$.

* ND, not determined. The concentration of the enantiomer product in these reactions was too low to obtain a reliable $\delta^{15}\text{N}$ value.

While recent AMS data indicate that some of the initial amino acid racemization dates reported for palaeo-indian remains from California were incorrect¹⁴, verification of the younger AMS dates (Holocene) will, to a large extent, depend on establishing the absolute indigeneity of the proteinaceous material isolated from the bones. Simply, any dating method, whether it be amino acid racemization or ^{14}C , is limited by the extent to which the origin(s) of the organic matter being analysed can be established.

Amino acid constituents of organisms commonly have distinct stable carbon and nitrogen isotopic compositions. Pillinger¹⁹ has suggested that the indigeneity of partially racemized or racemic amino acids in a sample might be established by measuring the stable carbon or nitrogen isotopic compositions of their respective D- and L-enantiomers. This could work because currently accepted mechanisms for the amino acid racemization reaction²⁰, such as abstraction of a proton from the α -carbon, imply that the individual amino acids should retain their stable nitrogen and carbon isotopic integrity during racemization. Subsequent input, for example, of non-indigenous L-amino acids, might, in some cases, be distinguished by differences in the stable isotope values of the D- and L-enantiomers.

An analytical technique has been designed for the isolation of D- and L-enantiomers of amino acids for stable isotopic analyses of glutamic acid²¹. Glutamic acid is a common protein amino acid that racemizes fairly rapidly in fossil bone²² and is a potential contaminant of fossil material through a bacterial contribution^{16,22}. Before attempting the difficult analyses of the stable isotopic compositions of individual D- and L-amino acid enantiomers in fossil materials, it is necessary to verify that amino acids will, in fact, retain their stable carbon and nitrogen isotopic integrity during racemization.

In this present study, laboratory experiments were conducted to evaluate the effects of racemization on the stable carbon and nitrogen isotopic compositions of amino acid enantiomers. Aliquots (1 ml) of a 0.01 M aqueous solution (pH 1.0) of L-glutamic acid (L-Glu; Nutritional Biochemical Corporation) were sealed in Pyrex tubes under N_2 and heated at 100 °C for up to 1,772 h. Additional aliquots were heated at 150 °C for up to 264 h. The experiments were repeated, substituting a 0.01 M aqueous solution of D-glutamic acid (D-Glu; NBC) for L-glutamic acid as the starting solution.

After heating, 50- μl aliquots were removed from each tube, evaporated to dryness in reaction vials, derivatized (to *N*-pentafluoropropionyl-isopropyl esters) and analysed for their respective D/L (or L/D) values by GC using a Chirasil-Val 30 m $\times$ 0.25 mm i.d. wall coated open tubular fused-silica column¹¹. Glutamic acid enantiomeric ratios were confirmed by analysing 50- μl aliquots of the heated samples by HPLC using an LC-18 reversed-phase column (25 cm $\times$ 4.6 mm i.d.; 5 μm particle-size resin) and an aqueous chiral mobile phase (isocratic; pH = 5.51) consisting of 0.008 M *N,N*-di-*n*-propyl-L-alanine and 0.004 M copper acetate²³. The D/L and L/D values are listed in Table 1. At 100 °C, the racemization of L-glutamic acid and D-glutamic acid followed approximately identical first-order reversible rate

laws ($k \approx 1.6 \times 10^{-4} \text{ h}^{-1}$). The samples heated at 150 °C were racemic after 264 h.

The separation and purification of D-glutamic acid and L-glutamic acid in the respective solutions was accomplished using a combination of HPLC and cation-exchange procedures²¹. The individual D-glutamic acid and L-glutamic acid sample residues were converted to gas and analysed for their stable carbon and nitrogen isotopic compositions by Dumas sealed-tube combustion in quartz^{24,25}. Stable nitrogen and carbon isotopic values for the reactants and products are listed in Table 2.

Whether the starting material is L-glutamic acid or D-glutamic acid, the resultant D- and L-enantiomer products retain their stable carbon isotopic integrity after 1,772 h of heating at 100 °C, as well as at equilibrium in the 150 °C heating experiment. While the $\delta^{15}\text{N}$ values remained essentially unchanged during heating experiments conducted at 100 °C, it appears that during heating at 150 °C, the enantiomer products became enriched in ^{15}N ; that is, the $\delta^{15}\text{N}$ values became more positive for the D- and L-enantiomer products. Note that when either D-glutamic acid or L-glutamic acid was heated for 264 h at 150 °C, the enantiomer product that had the reverse stereochemistry of the initial reactant (for example, the L-glutamic acid product, when the starting material was D-glutamic acid) was significantly more enriched in ^{15}N than the enantiomer product with the same stereochemistry as the initial reactant. It has been previously reported that glutamic acid is relatively stable at elevated temperatures²⁶. This stability is in part attributed to the facility by which glutamic acid in aqueous solution can undergo a reversible cyclization reaction to form pyroglutamic acid^{16,26}. However, given the acidic nature of the starting solution in the present study (pH = 1.0), cyclization of glutamic acid would not be favoured²⁶. While it is speculated, at present, that partial deamination of glutamic acid at 150 °C may have released ammonia that was depleted in ^{15}N , resulting in more positive $\delta^{15}\text{N}$ values of the enantiomer products, additional experiments will be required to elucidate the cause(s) of this observed shift.

The results of the present study indicate that, after the death of an organism, ensuing low-temperature, diagenetic reactions such as racemization should not cause a significant shift in the stable carbon isotopic compositions of the resultant enantiomer products. While the 150 °C experiments were anomalous with respect to the $\delta^{15}\text{N}$ values, the 100 °C experimental results indicate that, in less extreme thermal conditions, the stable nitrogen isotopic integrity of the D- and L-enantiomers of amino acids may also be preserved.

Today there are few, if any, reliable criteria for establishing the absolute indigeneity of amino acids in fossils. Without such criteria, geochronological, palaeoclimatic and ancient dietary pattern interpretations based on geochemical analyses of amino acid constituents of fossils will, in many cases, remain suspect. Our experimental results indicate that, in cases where sufficient organic matter is preserved in a fossil to permit the separation and stable isotopic analyses of the D- and L-enantiomers of an amino acid, this approach could provide a first step towards

delineating the absolute indigeneity of amino acids in fossils. As AMS techniques are refined to handle smaller samples, it may also become possible to date individual amino acid enantiomers by the ^{14}C method. If one enantiomer is entirely derived from the other by racemization during diagenesis, the individual D- and L-enantiomers for a given amino acid should have identical ^{14}C ages.

Older, more poorly preserved fossils may not always prove amenable to the determination of amino acid indigeneity by the stable isotope method, as the prospects for complete replacement of indigenous amino acids with non-indigenous amino acids increases with time. As non-indigenous amino acids undergo racemization, the enantiomers may have identical isotopic compositions and still not be related to the original organisms. Such a circumstance may, however, become easier to recognize as more information becomes available concerning the distribution and stable isotopic composition of the amino acid constituents of modern representatives of fossil organisms. Also, AMS dates on individual amino acid enantiomers may, in some cases, help to clarify indigeneity problems, in particular when stratigraphic controls can be used to estimate a general age range for the fossil in question.

Finally, the development of techniques for determining the stable isotopic composition of amino acid enantiomers may enable us to establish whether non-racemic amino acids in some carbonaceous meteorites²⁷ are indigenous, or result in part from terrestrial contamination.

M.H.E. thanks the NSF, Division of Earth Sciences (grant EAR-8352055) and the following contributors to his Presidential Young Investigator Award for partial support of this research:

Arco, Exxon, Phillips Petroleum, Texaco Inc., The Upjohn Co. We also acknowledge the donors of the Petroleum Research Fund, administered by the American Chemical Society (grant 16144-AC2 to M.H.E., grant 14805-AC2 to S.A.M.) for support. S.A.M. acknowledges NSERC (grant A2644) for partial support.

Received 19 May; accepted 15 July 1986.

1. Bada, J. L. & Protch, R. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1331-1334 (1973).
2. Bada, J. L., Schroeder, R. A. & Carter, G. F. *Science* **184**, 791-793 (1974).
3. Boulton, G. S. *et al. Nature* **298**, 437-441 (1982).
4. Wehmiller, J. F. in *Quaternary Dating Methods* (ed. Mahaney, W. C.) 171-193 (Elsevier, Amsterdam, 1984).
5. Engel, M. H., Zumberge, J. E. & Nagy, B. *Analyt. Biochem.* **82**, 415-422 (1977).
6. Bada, J. L. *A. Rev. Earth planet. Sci.* **13**, 241-268 (1985).
7. Chisholm, B. S., Nelson, D. E. & Schwarcz, H. P. *Science* **216**, 1131-1132 (1982).
8. Ambrose, S. H. & DeNiro, M. J. *Nature* **319**, 321-324 (1986).
9. Macko, S. A., Estep, M. L. F., Hare, P. E. & Hoering, T. C. *Yb. Carnegie Instn Wash.* **82**, 404-410 (1983).
10. Hare, P. E. & Estep, M. L. F. *Yb. Carnegie Instn Wash.* **82**, 410-414 (1983).
11. Engel, M. H. & Hare, P. E. in *Chemistry and Biochemistry of the Amino Acids* (ed. Barrett, G. C.) 462-479 (Chapman and Hall, London, 1985).
12. Johnstone, R. A. W. & Rose, M. E. in *Chemistry and Biochemistry of the Amino Acids* (ed. Barrett, G. C.) 480-524 (Chapman and Hall, London, 1985).
13. Weinstein, S., Engel, M. H. & Hare, P. E. in *Practical Protein Chemistry—A Handbook* (ed. Darbre, A.) 337-344 (Wiley, New York, 1986).
14. Bada, J. L., Gillespie, R., Gowllett, J. A. J. & Hedges, R. E. M. *Nature* **312**, 442-444 (1984).
15. Mitterer, R. M. & Kriausakul, N. *Org. Geochem.* **7**, 91-98 (1984).
16. Williams, K. M. & Smith, G. G. *Origins Life* **8**, 91-144 (1977).
17. Engel, M. H. & Hare, P. E. *Yb. Carnegie Instn Wash.* **81**, 425-430 (1982).
18. Hare, P. E. *Yb. Carnegie Instn Wash.* **73**, 576-581 (1974).
19. Pillinger, C. T. *Nature* **296**, 802 (1982).
20. Neuberger, A. *Adv. Protein Chem.* **4**, 298-383 (1948).
21. Engel, M. H. & Macko, S. A. *Analyt. Chem.* **56**, 2598-2600 (1984).
22. Dungworth, G. *Chem. Geol.* **17**, 135-153 (1976).
23. Weinstein, S., Engel, M. H. & Hare, P. E. *Analyt. Biochem.* **121**, 370-377 (1982).
24. Macko, S. A., Lee, W. Y. & Parker, P. L. *J. exp. mar. Biol. Ecol.* **63**, 145-149 (1982).
25. Macko, S. A., Estep, M. L. F. & Hoering, T. C. *Yb. Carnegie Instn Wash.* **81**, 413-417 (1982).
26. Vallentyne, J. R. *Geochim. cosmochim. Acta* **28**, 157-188 (1964).
27. Engel, M. H. & Nagy, B. *Nature* **296**, 837-840 (1982).

Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton†
& Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA

† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA

We describe a new learning procedure, back-propagation, for networks of neurone-like units. The procedure repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector. As a result of the weight adjustments, internal 'hidden' units which are not part of the input or output come to represent important features of the task domain, and the regularities in the task are captured by the interactions of these units. The ability to create useful new features distinguishes back-propagation from earlier, simpler methods such as the perceptron-convergence procedure¹.

There have been many attempts to design self-organizing neural networks. The aim is to find a powerful synaptic modification rule that will allow an arbitrarily connected neural network to develop an internal structure that is appropriate for a particular task domain. The task is specified by giving the desired state vector of the output units for each state vector of the input units. If the input units are directly connected to the output units it is relatively easy to find learning rules that iteratively adjust the relative strengths of the connections so as to progressively reduce the difference between the actual and desired output vectors². Learning becomes more interesting but

more difficult when we introduce hidden units whose actual or desired states are not specified by the task. (In perceptrons, there are 'feature analysers' between the input and output that are not true hidden units because their input connections are fixed by hand, so their states are completely determined by the input vector: they do not learn representations.) The learning procedure must decide under what circumstances the hidden units should be active in order to help achieve the desired input-output behaviour. This amounts to deciding what these units should represent. We demonstrate that a general purpose and relatively simple procedure is powerful enough to construct appropriate internal representations.

The simplest form of the learning procedure is for layered networks which have a layer of input units at the bottom; any number of intermediate layers; and a layer of output units at the top. Connections within a layer or from higher to lower layers are forbidden, but connections can skip intermediate layers. An input vector is presented to the network by setting the states of the input units. Then the states of the units in each layer are determined by applying equations (1) and (2) to the connections coming from lower layers. All units within a layer have their states set in parallel, but different layers have their states set sequentially, starting at the bottom and working upwards until the states of the output units are determined.

The total input, x_j , to unit j is a linear function of the outputs, y_i , of the units that are connected to j and of the weights, w_{ji} , on these connections

$$x_j = \sum_i y_i w_{ji} \quad (1)$$

Units can be given biases by introducing an extra input to each unit which always has a value of 1. The weight on this extra input is called the bias and is equivalent to a threshold of the opposite sign. It can be treated just like the other weights.

A unit has a real-valued output, y_j , which is a non-linear function of its total input

$$y_j = \frac{1}{1 + e^{-x_j}} \quad (2)$$

† To whom correspondence should be addressed.

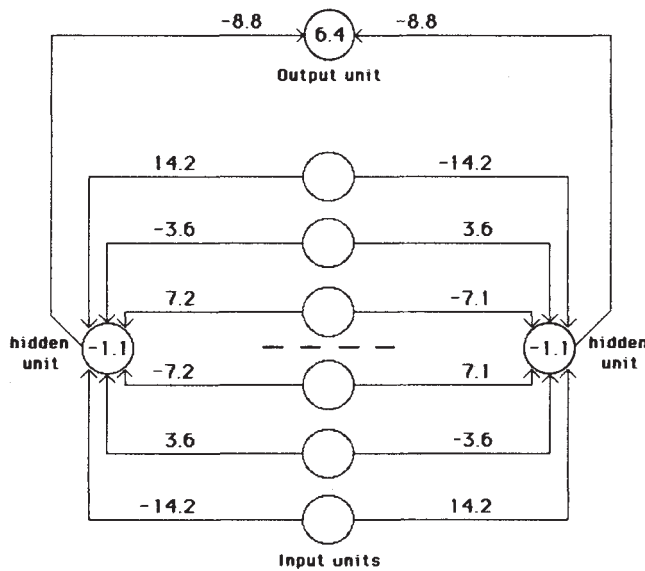


Fig. 1 A network that has learned to detect mirror symmetry in the input vector. The numbers on the arcs are weights and the numbers inside the nodes are biases. The learning required 1,425 sweeps through the set of 64 possible input vectors, with the weights being adjusted on the basis of the accumulated gradient after each sweep. The values of the parameters in equation (9) were $\epsilon = 0.1$ and $\alpha = 0.9$. The initial weights were random and were uniformly distributed between -0.3 and 0.3 . The key property of this solution is that for a given hidden unit, weights that are symmetric about the middle of the input vector are equal in magnitude and opposite in sign. So if a symmetrical pattern is presented, both hidden units will receive a net input of 0 from the input units, and, because the hidden units have a negative bias, both will be off. In this case the output unit, having a positive bias, will be on. Note that the weights on each side of the midpoint are in the ratio 1:2:4. This ensures that each of the eight patterns that can occur above the midpoint sends a unique activation sum to each hidden unit, so the only pattern below the midpoint that can exactly balance this sum is the symmetrical one. For all non-symmetrical patterns, both hidden units will receive non-zero activations from the input units. The two hidden units have identical patterns of weights but with opposite signs, so for every non-symmetric pattern one hidden unit will come on and suppress the output unit.

It is not necessary to use exactly the functions given in equations (1) and (2). Any input-output function which has a bounded derivative will do. However, the use of a linear function for combining the inputs to a unit before applying the nonlinearity greatly simplifies the learning procedure.

The aim is to find a set of weights that ensure that for each input vector the output vector produced by the network is the same as (or sufficiently close to) the desired output vector. If there is a fixed, finite set of input-output cases, the total error in the performance of the network with a particular set of weights can be computed by comparing the actual and desired output vectors for every case. The total error, E , is defined as

$$E = \frac{1}{2} \sum_c \sum_j (y_{j,c} - d_{j,c})^2 \quad (3)$$

where c is an index over cases (input-output pairs), j is an index over output units, y is the actual state of an output unit and d is its desired state. To minimize E by gradient descent it is necessary to compute the partial derivative of E with respect to each weight in the network. This is simply the sum of the partial derivatives for each of the input-output cases. For a given case, the partial derivatives of the error with respect to each weight are computed in two passes. We have already described the forward pass in which the units in each layer have their states determined by the input they receive from units in lower layers using equations (1) and (2). The backward pass which propagates derivatives from the top layer back to the bottom one is more complicated.

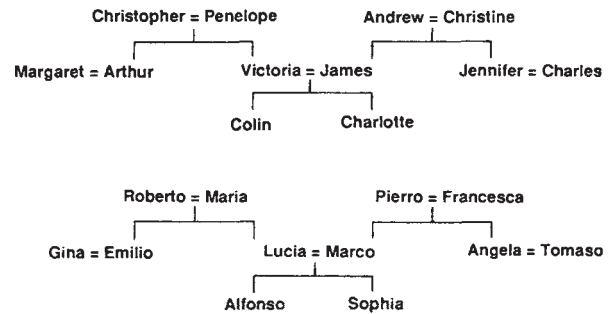


Fig. 2 Two isomorphic family trees. The information can be expressed as a set of triples of the form (person 1)(relationship)(person 2), where the possible relationships are {father, mother, husband, wife, son, daughter, uncle, aunt, brother, sister, nephew, niece}. A layered net can be said to 'know' these triples if it can produce the third term of each triple when given the first two. The first two terms are encoded by activating two of the input units, and the network must then complete the proposition by activating the output unit that represents the third term.

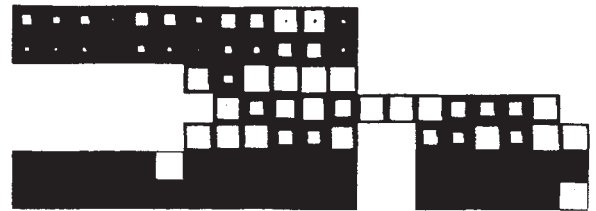


Fig. 3 Activity levels in a five-layer network after it has learned. The bottom layer has 24 input units on the left for representing (person 1) and 12 input units on the right for representing the relationship. The white squares inside these two groups show the activity levels of the units. There is one active unit in the first group representing Colin and one in the second group representing the relationship 'has-aunt'. Each of the two input groups is totally connected to its own group of 6 units in the second layer. These groups learn to encode people and relationships as distributed patterns of activity. The second layer is totally connected to the central layer of 12 units, and these are connected to the penultimate layer of 6 units. The activity in the penultimate layer must activate the correct output units, each of which stands for a particular (person 2). In this case, there are two correct answers (marked by black dots) because Colin has two aunts. Both the input units and the output units are laid out spatially with the English people in one row and the isomorphic Italians immediately below.

The backward pass starts by computing $\partial E / \partial y_j$ for each of the output units. Differentiating equation (3) for a particular case, c , and suppressing the index c gives

$$\partial E / \partial y_j = y_j - d_j \quad (4)$$

We can then apply the chain rule to compute $\partial E / \partial x_j$

$$\partial E / \partial x_j = \partial E / \partial y_j \cdot dy_j / dx_j$$

Differentiating equation (2) to get the value of dy_j / dx_j and substituting gives

$$\partial E / \partial x_j = \partial E / \partial y_j \cdot y_j (1 - y_j) \quad (5)$$

This means that we know how a change in the total input x to an output unit will affect the error. But this total input is just a linear function of the states of the lower level units and it is also a linear function of the weights on the connections, so it is easy to compute how the error will be affected by changing these states and weights. For a weight w_{ji} , from i to j the derivative is

$$\begin{aligned} \partial E / \partial w_{ji} &= \partial E / \partial x_j \cdot \partial x_j / \partial w_{ji} \\ &= \partial E / \partial x_j \cdot y_i \end{aligned} \quad (6)$$

and for the output of the i^{th} unit the contribution to $\partial E / \partial y_i$

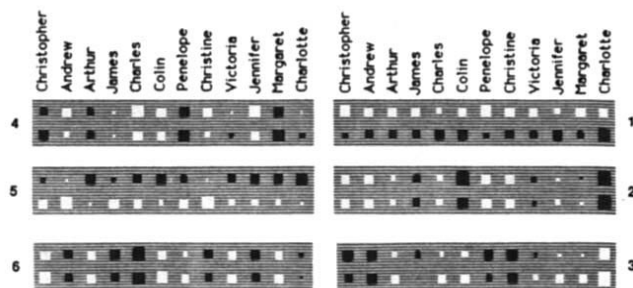


Fig. 4 The weights from the 24 input units that represent people to the 6 units in the second layer that learn distributed representations of people. White rectangles, excitatory weights; black rectangles, inhibitory weights; area of the rectangle encodes the magnitude of the weight. The weights from the 12 English people are in the top row of each unit. Unit 1 is primarily concerned with the distinction between English and Italian and most of the other units ignore this distinction. This means that the representation of an English person is very similar to the representation of their Italian equivalent. The network is making use of the isomorphism between the two family trees to allow it to share structure and it will therefore tend to generalize sensibly from one tree to the other. Unit 2 encodes which generation a person belongs to, and unit 6 encodes which branch of the family they come from. The features captured by the hidden units are not at all explicit in the input and output encodings, since these use a separate unit for each person. Because the hidden features capture the underlying structure of the task domain, the network generalizes correctly to the four triples on which it was not trained. We trained the network for 1500 sweeps, using $\epsilon = 0.005$ and $\alpha = 0.5$ for the first 20 sweeps and $\epsilon = 0.01$ and $\alpha = 0.9$ for the remaining sweeps. To make it easier to interpret the weights we introduced 'weight-decay' by decrementing every weight by 0.2% after each weight change. After prolonged learning, the decay was balanced by $\partial E/\partial w$, so the final magnitude of each weight indicates its usefulness in reducing the error. To prevent the network needing large weights to drive the outputs to 1 or 0, the error was considered to be zero if output units that should be on had activities above 0.8 and output units that should be off had activities below 0.2.

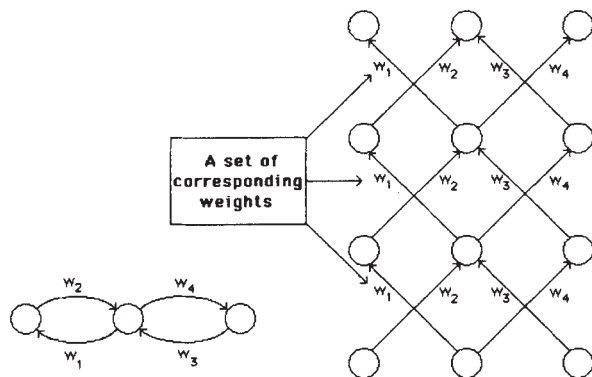


Fig. 5 A synchronous iterative net that is run for three iterations and the equivalent layered net. Each time-step in the recurrent net corresponds to a layer in the layered net. The learning procedure for layered nets can be mapped into a learning procedure for iterative nets. Two complications arise in performing this mapping: first, in a layered net the output levels of the units in the intermediate layers during the forward pass are required for performing the backward pass (see equations (5) and (6)). So in an iterative net it is necessary to store the history of output states of each unit. Second, for a layered net to be equivalent to an iterative net, corresponding weights between different layers must have the same value. To preserve this property, we average $\partial E/\partial w$ for all the weights in each set of corresponding weights and then change each weight in the set by an amount proportional to this average gradient. With these two provisos, the learning procedure can be applied directly to iterative nets. These nets can then either learn to perform iterative searches or learn sequential structures⁴.

resulting from the effect of i on j is simply

$$\partial E/\partial x_j \cdot \partial x_j/\partial y_i = \partial E/\partial x_j \cdot w_{ji}$$

so taking into account all the connections emanating from unit i we have

$$\partial E/\partial y_i = \sum_j \partial E/\partial x_j \cdot w_{ji} \quad (7)$$

We have now seen how to compute $\partial E/\partial y$ for any unit in the penultimate layer when given $\partial E/\partial y$ for all units in the last layer. We can therefore repeat this procedure to compute this term for successively earlier layers, computing $\partial E/\partial w$ for the weights as we go.

One way of using $\partial E/\partial w$ is to change the weights after every input-output case. This has the advantage that no separate memory is required for the derivatives. An alternative scheme, which we used in the research reported here, is to accumulate $\partial E/\partial w$ over all the input-output cases before changing the weights. The simplest version of gradient descent is to change each weight by an amount proportional to the accumulated $\partial E/\partial w$

$$\Delta w = -\epsilon \partial E/\partial w \quad (8)$$

This method does not converge as rapidly as methods which make use of the second derivatives, but it is much simpler and can easily be implemented by local computations in parallel hardware. It can be significantly improved, without sacrificing the simplicity and locality, by using an acceleration method in which the current gradient is used to modify the velocity of the point in weight space instead of its position

$$\Delta w(t) = -\epsilon \partial E/\partial w(t) + \alpha \Delta w(t-1) \quad (9)$$

where t is incremented by 1 for each sweep through the whole set of input-output cases, and α is an exponential decay factor between 0 and 1 that determines the relative contribution of the current gradient and earlier gradients to the weight change.

To break symmetry we start with small random weights. Variants on the learning procedure have been discovered independently by David Parker (personal communication) and by Yann Le Cun³.

One simple task that cannot be done by just connecting the input units to the output units is the detection of symmetry. To detect whether the binary activity levels of a one-dimensional array of input units are symmetrical about the centre point, it is essential to use an intermediate layer because the activity in an individual input unit, considered alone, provides no evidence about the symmetry or non-symmetry of the whole input vector, so simply adding up the evidence from the individual input units is insufficient. (A more formal proof that intermediate units are required is given in ref. 2.) The learning procedure discovered an elegant solution using just two intermediate units, as shown in Fig. 1.

Another interesting task is to store the information in the two family trees (Fig. 2). Figure 3 shows the network we used, and Fig. 4 shows the 'receptive fields' of some of the hidden units after the network was trained on 100 of the 104 possible triples.

So far, we have only dealt with layered, feed-forward networks. The equivalence between layered networks and recurrent networks that are run iteratively is shown in Fig. 5.

The most obvious drawback of the learning procedure is that the error-surface may contain local minima so that gradient descent is not guaranteed to find a global minimum. However, experience with many tasks shows that the network very rarely gets stuck in poor local minima that are significantly worse than the global minimum. We have only encountered this undesirable behaviour in networks that have just enough connections to perform the task. Adding a few more connections creates extra dimensions in weight-space and these dimensions provide paths around the barriers that create poor local minima in the lower dimensional subspaces.

The learning procedure, in its current form, is not a plausible model of learning in brains. However, applying the procedure to various tasks shows that interesting internal representations can be constructed by gradient descent in weight-space, and this suggests that it is worth looking for more biologically plausible ways of doing gradient descent in neural networks.

We thank the System Development Foundation and the Office of Naval Research for financial support.

Received 1 May; accepted 31 July 1986.

1. Rosenblatt, F. *Principles of Neurodynamics* (Spartan, Washington, DC, 1961).
2. Minsky, M. L. & Papert, S. *Perceptrons* (MIT, Cambridge, 1969).
3. Le Cun, Y. *Proc. Cognitiva* 85, 599-604 (1985).
4. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1: *Foundations* (eds Rumelhart, D. E. & McClelland, J. L.) 318-362 (MIT, Cambridge, 1986).

Bilateral amblyopia after a short period of reverse occlusion in kittens

Kathryn M. Murphy* & Donald E. Mitchell

Department of Psychology, Dalhousie University,
Halifax Nova Scotia, Canada B3H 4J1

The majority of neurones in the visual cortex of both adult cats and kittens can be excited by visual stimulation of either eye. Nevertheless, if one eye is deprived of patterned vision early in life, most cortical cells can only be activated by visual stimuli presented to the nondeprived eye and behaviourally the deprived eye is apparently useless^{1,2}. Although the consequences of monocular deprivation can be severe, they can in many circumstances be rapidly reversed with the early implementation of reverse occlusion which forces the use of the initially deprived eye^{3,4}. However, by itself reverse occlusion does not restore a normal distribution of cortical ocular dominance³ and only promotes visual recovery in one eye^{5,6}. In an effort to find a procedure that might restore good binocular vision, we have examined the effects on acuity and cortical ocular dominance of a short, but physiologically optimal period of reverse occlusion, followed by a period of binocular vision beginning at 7.5 weeks of age. Surprisingly, despite the early introduction of binocular vision, both eyes attained acuities that were only approximately 1/3 of normal acuity levels. Despite the severe bilateral amblyopia, cortical ocular dominance appeared similar to that of normal cats. This is the first demonstration of severe bilateral amblyopia following consecutive periods of monocular occlusion.

Nine kittens were used, of which eight were monocularly deprived by eyelid suture from about the time of natural eye opening (6 to 11 days) until 5 weeks of age, at which time the initially deprived eye was opened and the other eye was sutured closed for 18 days. Physiological recordings from area 17 were made from one normal control and from five monocularly-deprived kittens, one immediately after reverse occlusion (as a control); the remaining four after a further 4 weeks at least (range 4-8 weeks) of normal binocular vision. Grating acuity thresholds were determined for both eyes of a further three kittens (subjected to the same regime—monocular deprivation, 18 days reverse suturing, followed by normal binocular vision) by use of a jumping stand^{5,7}. None of the kittens tested behaviourally were examined physiologically. Single unit recordings were made in area 17 of the anaesthetized, paralysed kittens (one normal, five experimental) with glass coated platinum-iridium electrodes. Anaesthesia was induced by

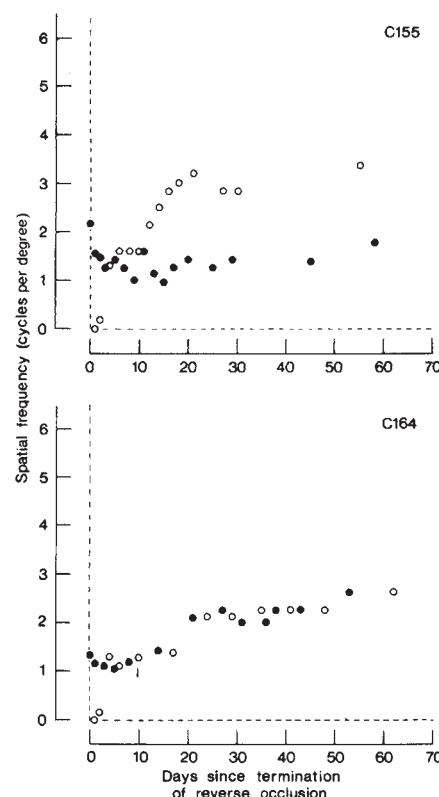


Fig. 1. Changes in visual acuity during the period of binocular vision for two kittens (C155 and C164) that were previously monocularly deprived until 5 weeks of age, and then reverse occluded for 18 days. ●, Acuity of the initially deprived eye; ○, acuity of the initially nondeprived eye.

intravenous pentothal and maintained by artificial respiration with 70% N₂O and 30% O₂ supplemented with intravenous Nembutal; EEG, EKG, body temperature, and expired CO₂ levels were monitored. The eyes were brought to focus on a tangent screen 137 cm distant from the kitten using contact lenses with 3 mm artificial pupils. Single units were recorded along one long penetration in area 17 down the medial bank of the postlateral gyrus in each hemisphere, always beginning in the hemisphere contralateral to the initially open eye. Receptive fields were sampled according to established procedures⁸, every 100 μ m along the penetration in a cortical region corresponding to the horizontal meridian of visual space. All units were located within 15° of the area centralis, with the majority within 5°.

The longitudinal changes in visual acuity of both eyes following introduction of binocular vision are shown in Fig. 1 for two representative kittens. At the end of 18 days of reverse occlusion the vision of the initially deprived eye had recovered to only rudimentary levels (1-2.5 cycles per degree) while at the same time the initially nondeprived eye had been rendered blind. During the subsequent period of binocular visual exposure the vision of both eyes improved slightly, but only to a very limited extent (to between 1.7 and 3.4 cycles per degree). The results from the third animal were very similar. After more than 2 months of binocular exposure the acuities of the initially deprived and nondeprived eyes were respectively, 2.54 and 3.35 cycles per degree. Surprisingly, after 2 months of binocular vision, the acuity of both eyes of these animals remained at about one-third to one-half of normal levels⁶. Although the initially deprived eye was opened at the peak of the sensitive period (5 weeks of age) and the initially nondeprived eye was closed for a relatively brief period of time (18 days), this deprivation regimen had a devastating and permanent effect upon the visual acuity of both eyes.

* Present address: School of Optometry, University of California, Berkeley, California 94720, USA.

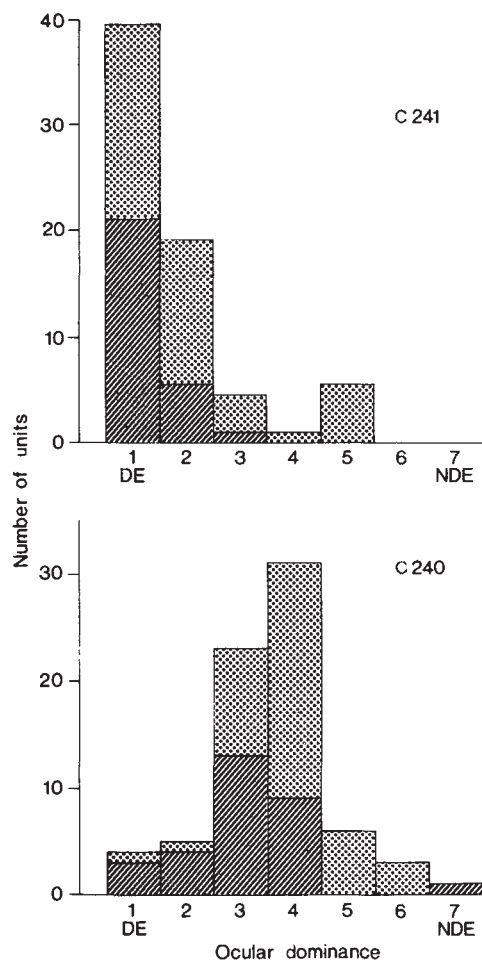


Fig. 2. Distribution of ocular dominance for one kitten that did not receive binocular visual experience following reverse occlusion (C241) and one kitten that subsequently received 4 weeks of binocular visual exposure (C240). The stippled and striped regions correspond to cells recorded from, respectively, the hemisphere contralateral to the initially nondeprived eye (NDE) and contralateral to the initially deprived eye (DE).

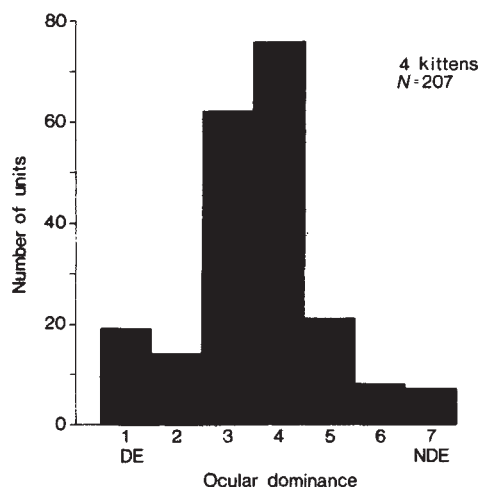


Fig. 3 The ocular dominance histogram for 207 cells recorded from four kittens that received binocular visual experience following reverse occlusion. The ocular dominance distribution has been plotted as if all recordings were made from the hemisphere contralateral to the initially deprived eye (DE).

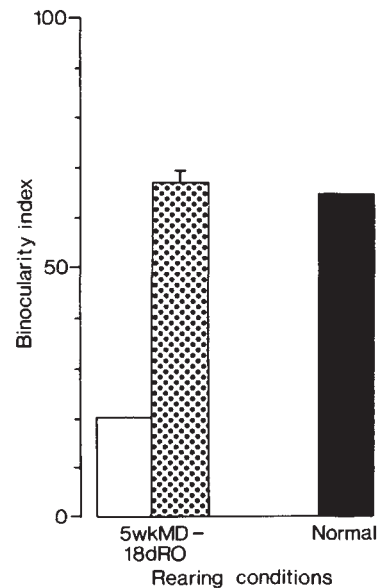


Fig. 4 The binocularity indices for one kitten that did not receive binocular vision (open column), the mean for four kittens that received binocular experience following reverse occlusion (stippled column), and a normal cat (dark column). The binocularity index was calculated as follows:

$$[(4 + 2/3(3 + 5) + 1/3(2 + 6)) / \text{total number of cells}] \times 100$$

where the italicized numbers represent the number of cells classified in the corresponding ocular dominance group. The number of cells in each group is weighted according to degree of dominance so that the index reflects the nature of ocular dominance distribution.

The ocular dominance distributions for the kitten from which recordings were made immediately upon termination of reverse occlusion (C241) and a kitten that subsequently received 4 weeks of binocular visual experience (C240) are shown in Fig. 2. Following 18 days of reverse occlusion there was an almost complete shift of cortical ocular dominance towards the initially deprived eye, with 90% of the units sampled dominated by this eye and only 9% of the units dominated by the initially nondeprived eye (Fig. 2, C241). These results demonstrate a degree of ocular dominance shift comparable to those reported by Movshon⁴. However, these results are quite different from those observed in the group of four kittens that received a period of binocular vision following reverse occlusion. As can be seen in both the ocular dominance histogram from one representative kitten from this group (C240) shown in Fig. 2, and also the composite histogram for data from all four kittens (Fig. 3), cells in area 17 were predominantly binocular following binocular visual experience. Qualitative examination of the receptive fields demonstrated the normal range of properties. An index of cortical binocularity was calculated in order to enable a quantitative comparison between the distribution of binocular units recorded either before or after binocular visual experience. This index has a value of 100 if all of the cells were equally activated through either eye (ocular dominance group 4), and a value of zero if all of the cells were exclusively monocular (that is, groups 1 and/or 7). Figure 4 shows the value of the index for a normal cat (dark bar), the kitten recorded from immediately upon termination of reverse occlusion (open bar) and the four kittens that had binocular experience (stippled bar). Interestingly, the latter kittens possessed a degree of cortical binocularity (index = 67) comparable to that of the normal adult cat.

Although a virtually normal degree of cortical binocularity was restored following binocular visual experience, this same deprivation regimen resulted in severe bilateral amblyopia.

These results demonstrate a very obvious discrepancy between physiological and behavioural measures. Simply increasing the efficacy of the cortical response to visual stimulation through an eye does not necessarily lead to a recovery of visual acuity. As postulated earlier⁹, the spatial resolution of the most sensitive neurones, rather than the number of units responsive to each eye, may be a better predictor of the behavioural visual acuity. However, the present physiological results show that both eyes make functional cortical connections which sustain receptive fields that exhibit qualitatively normal properties.

It is a puzzle why the cortical ocular dominance appears normal when the acuities are so poor. The poor acuity of both eyes along with the highly binocular ocular dominance distribution, make the development of a strabismus an unlikely explanation. One possible explanation for these results lies in a change in the organization of the geniculocortical afferents in layer 4 of area 17. Although previous work in monkeys has shown that reverse occlusion promotes expansion of the projections from the initially deprived eye^{10,11}, there have not been any detailed examinations of the changes to the initially nondeprived eye's projections during reverse occlusion. However, results of lateral geniculate nucleus (LGN) cell size measurements following reverse occlusion in monkeys indicate that the LGN cells receiving input from the initially nondeprived eye remain hyper-

trophied during reverse occlusion¹² and therefore may also be supporting an enlarged axonal arborization. Overly exuberant projections to the visual cortex from both eyes could lead to larger receptive fields and overlapping ocular dominance columns resulting in a smearing of the cortical retinotopic map. This smearing could generate a highly binocular cortex and at the same time lead to poor visual acuity through both eyes.

This work was supported by a grant from NSERC to D.E.M. (AP-7660) and a grant from the Medical Research Council of Canada (PG-29). K.M.M. held the R.C. Purse Fellowship from the Canadian National Institute for the Blind. We thank Dr M. Cynader for many helpful discussions and Dr N. V. Swindale for his comments on the manuscript.

Received 6 March; accepted 28 July 1986.

1. Wiesel, T. N. & Hubel, D. H. *J. Neurophysiol.* **26**, 1003-1017 (1963).
2. Wiesel, T. N. & Hubel, D. H. *J. Neurophysiol.* **28**, 1029-1040 (1965).
3. Blakemore, C. & Van Sluyters, R. C. *J. Physiol., Lond.* **248**, 663-716 (1975).
4. Movshon, J. A. *J. Physiol., Lond.* **261**, 125-174 (1976).
5. Mitchell, D. E., Murphy, K. M. & Kaye, M. *Invest. Ophthalmol. Vis. Sci.* **25**, 908-917 (1984).
6. Giffin, F. & Mitchell, D. E. *J. Physiol., Lond.* **274**, 511-537 (1978).
7. Mitchell, D. E., Giffin, F. & Timney, B. *Perception* **6**, 181-193 (1977).
8. Stryker, M. P. & Sherk, H. *Science* **190**, 904-905 (1975).
9. Mitchell, D. E., Cynader, M. & Movshon, J. A. *J. comp. Neurol.* **176**, 53-64 (1977).
10. Le Vay, S., Wiesel, T. N. & Hubel, D. H. *J. comp. Neurol.* **191**, 1-51 (1980).
11. Swindale, N. V., Vital-Durand, F. & Blakemore, C. *Proc. R. Soc. B213*, 435-450 (1981).
12. Sloper, J. J., Headon, M. P. & Powell, T. P. S. *Dev. Brain Res.* **15**, 295-297 (1984).

Chemotropic effect of specific target epithelium in the developing mammalian nervous system

A. G. S. Lumsden & Alun M. Davies

Department of Anatomy, Guy's Hospital Medical School, London SE1 9RT, UK, and St George's Hospital Medical School, London SW17, UK

Developing nerve fibres are guided to their targets by specific directional cues^{1,2} which are thought to be expressed in the tissues along the route³⁻⁷ and may involve the extracellular matrix⁸⁻¹⁰. Another possibility, that directional cues emanate from the target itself¹¹, is consistent with the recent demonstration of homing behaviour by ectopic retinal ganglion axons¹² and our previous demonstration that early trigeminal neurites grow directly to their virgin peripheral target *in vitro*¹³. Here we show that this chemotropic effect is precisely limited to the trigeminal system; trigeminal ganglion neurites grow directly to their own target field but not to the adjoining field, normally innervated by the geniculate ganglion; furthermore, the trigeminal field does not influence the growth of geniculate neurites. Also, when trigeminal ganglia are co-cultured with isolated tissue layers of their target, neurites grow only towards the epithelial and not the mesenchymal component. These findings suggest that trigeminal epithelium is specified to attract correct innervation and that pathway mesenchyme, in which preformed guidance cues have been postulated^{10,14,15}, may provide favourable conditions for nerve fibre growth but not govern its direction.

In the mouse embryo, nerve fibres first emerge from the trigeminal ganglion at E9.5 (day of vaginal plug is E0), and first contact the epithelium of the maxillary process (part of the peripheral target field of the trigeminal ganglion) during E11^{16,17}. The adult maxillary nerve consists of sensory fibres, many of which terminate in the epithelium as disk endings associated with Merkel cells in the external root sheaths of whiskers¹⁸. We have previously shown that maxillary processes explanted at E10 and E11 release an agent, immunochemically distinct from nerve growth factor (NGF)¹³ and laminin¹⁹, which elicits and directs neurite outgrowth from trigeminal ganglia of the same age co-cultured in a collagen gel. To understand the role of this

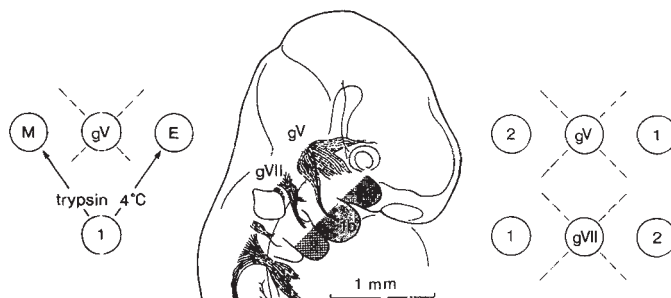
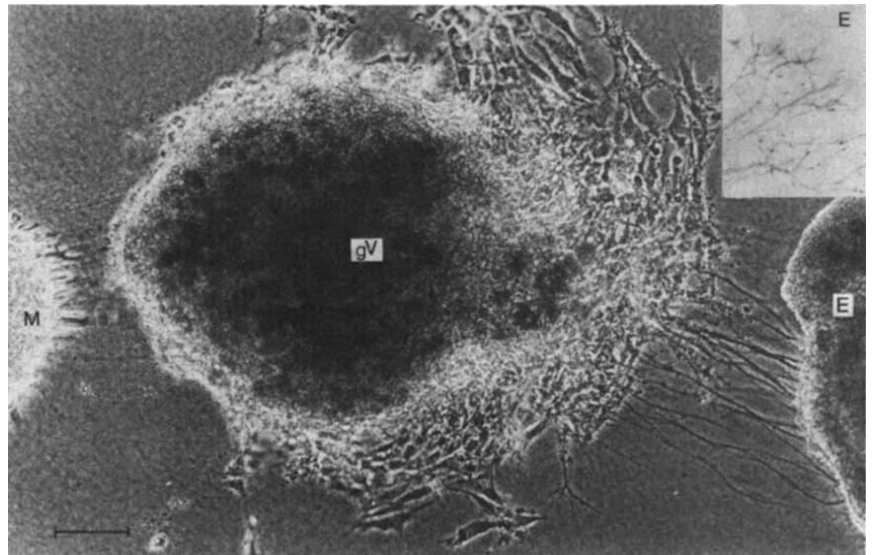


Fig. 1 Schematic representation of procedures in which (right) cranial sensory ganglia (gV, trigeminal, innervating the first branchial arch; gVII, geniculate, innervating the second branchial arch) of E10 (depicted to scale) and E11 CD1 mouse embryos were co-cultured with both their own and adjoining peripheral target field tissues (1a and 1b, maxillary and mandibular processes of first arch; 2, hyoid process of the second arch). In another set of experiments (left) trigeminal ganglia were co-cultured with the epithelial (E) and mesenchymal (M) component layers of the maxillary process (1a). These were separated by treatment with 1% crude trypsin (Gibco) in Ca- and Mg-free phosphate-buffered saline for 45 min at 4°C. Explants were positioned 0.1-0.4 mm apart in the combinations shown and suspended in 50-μl drops of collagen solution (rat tail collagen, 3 mg ml⁻¹; for method see ref. 13). Each collagen drop was overlain by 250 μl F14 medium (Imperial Labs), containing fetal bovine serum (Sera Lab) and sheep anti-mouse NGF antiserum (final concentrations 10% and 0.2% respectively). After 24, 48 and 72 h incubation, neurites were visualized by phase contrast microscopy and by indirect immunoperoxidase using a monoclonal antibody (RT97) to neurofilaments. At 48 h the number, length and orientation of neurites in each ganglion quadrant (marked in the figure by dashed lines) were measured (see ref. 13).

agent and the mechanism by which it acts, it is necessary to establish, first, whether there is regional specificity in its production and in the neuronal populations which respond, and second, the identity of the cells which synthesize it.

The specificity of the agent was investigated by co-culturing the trigeminal ganglion with part of its cutaneous target field (either the maxillary or the mandibular process of the first branchial arch) and the adjoining cutaneous target field of the geniculate ganglion (the hyoid process of the second branchial arch), and by co-culturing the geniculate ganglion with both its

Fig. 2 Phase contrast photomicrograph of a representative collagen gel co-culture of an E10 mouse embryo trigeminal ganglion (gV) and E10 maxillary process mesenchyme (M) and epithelium (E). Neurites have extended directly to the epithelium but none has extended laterally or towards the mesenchyme. In this particular culture, collagen fibrils have become aligned on the mesenchyme side of the ganglion, presumably through cell traction, but not on the epithelium side. Non-neuronal cell outgrowth did not correlate with neurite outgrowth. The inset shows E10 trigeminal ganglion neurites growing directly to a monolayer of maxillary process epithelium (E); neurites are stained by the indirect immunoperoxidase method using anti-neurofilament monoclonal RT 97. Scale bar, 100 μ m.



own and trigeminal targets (Fig. 1, right). In all trigeminal ganglion co-cultures in which neurite outgrowth was observed (54% at E10, $n = 85$; 63% at E11, $n = 65$), the neurites were confined to the quadrant facing the first arch explant; in no co-cultures were neurites seen in the quadrant facing the second arch explant. Trigeminal ganglia cultured alone in a collagen gel did not extend neurites. Geniculate ganglia, however, extended neurites when cultured in isolation at both E10 and E11, and this outgrowth was neither augmented by nor oriented towards co-cultured target tissue of either arch.

The identity of the cells which synthesize the agent was investigated by co-culturing trigeminal ganglia with the epithelial and mesenchymal components of the maxillary process (Fig. 1, left). After 48 h neurite outgrowth was observed in 15% of E10 co-cultures ($n = 160$) and in 27.8% of E11 co-cultures ($n = 90$); in all cases neurites were present only in the quadrant facing the epithelium, and no neurites were present in the quadrant facing the mesenchyme or in lateral quadrants (Fig. 2). Neurites first appeared at 24 h and by 48 h the majority were close to or in contact with epithelium.

Epithelium-directed neurite outgrowth was enhanced in cases where epithelium had attached and spread on the plastic substratum beneath the collagen matrix. For this reason, we set up co-cultures in which the mesenchymal masses and epithelial sheets were first allowed to adhere to the base of the dish for a period of 4 h before being overlain with collagen solution in which ganglia were positioned as shown in Fig. 1. In these experiments, epithelium-directed outgrowth occurred in 60% of cultures ($n = 20$) and, as before, there was no neurite outgrowth towards mesenchyme. Outgrowth in these cultures was generally more profuse (Fig. 2 inset) than that which grew towards epithelium explants suspended in the collagen matrix.

Developing nerve fibres may be guided to their targets by two general mechanisms: either they respond to cues laid down along the route to the target, as has been suggested in some systems³⁻⁷, or they respond by chemotaxis to a long-range, target-derived signal¹¹. In a given system either one or both mechanisms could be involved. Our findings that early trigeminal but not geniculate neurites respond by directional growth to an agent whose production is localized to the trigeminal territory and which emanates from the very peripheral region of this territory (epithelium) is clear evidence for the operation of the second of these mechanisms.

This evidence raises the issue of whether the fibres of other cranial sensory ganglia and dorsal root ganglia (DRG) are guided by a similar mechanism. In addition to the lack of directional influence of the hyoid process on geniculate neurites in culture, the trigeminal system displays some distinctive features which indicate that it may be a special case. The receptive field of the trigeminal ganglion is the most densely innervated region of the periphery yet it is precisely circumscribed, in contrast to the considerable overlap between the receptive fields of DRG. Furthermore, in spinal nerves, motor axon outgrowth from the cord precedes outgrowth from DRG and has been proposed to provide a guidance pathway for sensory fibre outgrowth²⁰. In the trigeminal nerve, however, such stereotopic cues are not available; the major division (maxillary) is exclusively sensory and the mandibular division includes motor fibres for only a short proximal part of its course to the periphery.

We thank Yves Barde for the gift of anti-NGF antiserum, John Wood for the gift of RT 97 antibody, and Vanessa Harrison and Martin Sadler for technical assistance. This work was supported by a project grant from the MRC.

Received 7 April; accepted 16 July 1986.

1. Lance-Jones, C. & Landmesser, L. *Proc. R. Soc. Lond.* **B214**, 19-52 (1981).
2. Eisen, J. S., Myers, P. Z. & Westerfield, M. *Nature* **320**, 269-271 (1986).
3. Tosney, K. W. & Landmesser, L. *J. Neurosci.* **4**, 2518-2527 (1984).
4. Bonhoeffer, F. & Huf, J. *Nature* **288**, 162-164 (1980).
5. Silver, J. & Ogawa, M. Y. *Science* **220**, 1067-1069 (1983).
6. Keynes, R. J. & Stern, C. D. *Nature* **310**, 786-789 (1984).
7. Easter, S. S., Bratton, B. & Scherer, S. S. *J. Neurosci.* **4**, 2173-2190 (1984).
8. Lander, A. D., Fujii, D. K. & Reichardt, L. F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2183-2187 (1985).
9. Davis, G. E., Manthorpe, M., Engvall, E. & Varon, S. *J. Neurosci.* **5**, 2662-2671 (1985).

10. Rogers, S. L., Edson, K. J., Letourneau, P. C. & McLoon, S. C. *Dev. Biol.* **113**, 429-435 (1986).
11. Cajal, S. R. *Trab. Lab. invest. Biol. Univ. Madr.* **17**, 181-228 (1919).
12. Harris, W. *Nature* **320**, 266-269 (1986).
13. Lumsden, A. G. S. & Davies, A. M. *Nature* **306**, 786-788 (1983).
14. Ebendal, T. *Cell Tissue Res.* **175**, 439-458 (1977).
15. Al-Ghaith, L. K. & Lewis, J. H. *J. Embryol. exp. Morph.* **68**, 149-160 (1982).
16. Davies, A. M. & Lumsden, A. G. S. *J. comp. Neurol.* **223**, 124-137 (1984).
17. Davies, A. M. & Lumsden, A. G. S. *J. comp. Neurol.* (in the press).
18. Patrizi, G. & Munger, B. L. *J. comp. Neurol.* **126**, 423-436 (1966).
19. Lumsden, A. G. S. & Davies, A. M. in *Epithelial-Mesenchymal Interactions in Neural Development*, NATO-ASI Ser. (eds Wolff, J. R. & Sievers, J.) (Springer, New York, in the press).
20. Tosney, K. W. & Landmesser, L. *J. Neurosci.* **5**, 2345-2358 (1985).

A novel family of human major histocompatibility complex-related genes not mapping to chromosome 6

F. Calabi & C. Milstein

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Thymocyte antigens CD1 [Thy, gp45, 12] are thought to be the human counterparts of mouse thymus leukaemia (TL) antigens^{1,2}. Serological and biochemical analyses indicate that at least three subsets exist³⁻⁶, the first of which (HTA 1/T6) was initially identified by the monoclonal antibody NA1/34 (refs 7, 8). Like TL, CD1 are expressed on cortical thymocytes as well as on some lymphoid neoplasias, and resemble in structure major histocompatibility complex (MHC) class I antigens. However HTA 1/T6 is loosely associated with β_2 -microglobulin⁹⁻¹⁰ and is also found linked by a disulphide bridge to CD8 (T8)^{11,12}. A molecular genetic approach is needed to investigate the CD1 system, to clarify its relationship to TL antigens and to understand its regulation. We report the isolation of complementary DNA (cDNA) clones encoding a CD1 antigen. These clones reveal a novel family of genes which are MHC-related but are neither equivalent to mouse TL antigens nor linked to the MHC.

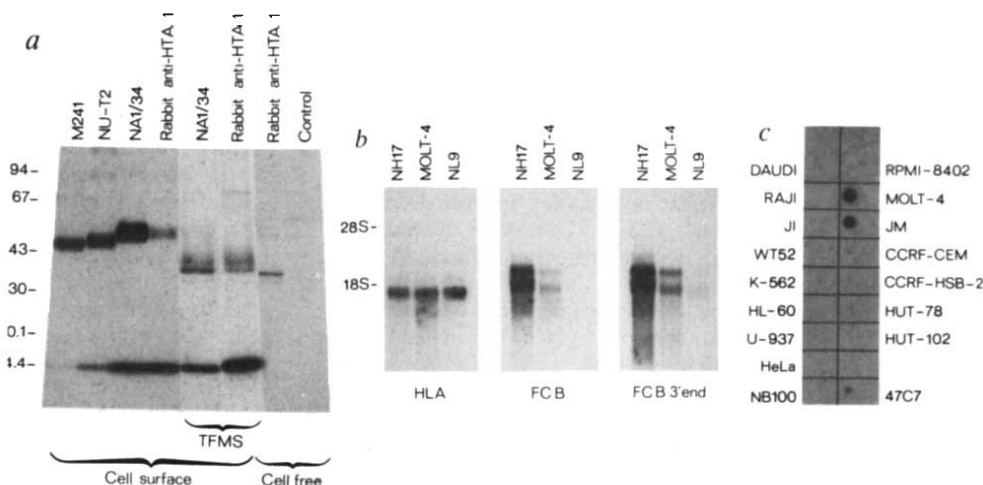
The isolation of CD1 cDNA clones involved derivation of a

cell line that expresses high levels of HTA 1, enrichment of HTA 1 mRNA by immunoselection of polysomes and screening of cDNA clones by mRNA selection followed by *in vitro* translation and immunoprecipitation. NH17 is a high HTA 1 expression mutant of MOLT-4¹³. A rabbit antiserum raised against HTA 1 purified from this cell line (ref. 10 and Kefford, R. F. & Calabi, F., unpublished) precipitates a β_2 -microglobulin-associated 49 kDa α chain from surface-iodinated NH17. This differs from the CD1 α chains recognized by monoclonal antibodies M241 and NU-T2 but comigrates with HTA 1 even after deglycosylation by trifluoromethanesulphonic (TFMS) acid treatment. Unlike NA1/34, the rabbit antiserum also recognizes the HTA 1 α chain synthesized *in vitro* (fig. 1a). As described in the legend to Fig. 1, a preparation enriched in HTA 1 mRNA was copied into cDNA and cloned into M13mp11¹⁶. One clone (FCB6) was found to select a mRNA encoding an HTA 1 α chain-like polypeptide according to relative molecular mass and serum specificity. The insert was later used to probe a total NH17 cDNA library made in λ gt11 (ref. 18) to yield eight cross-hybridizing clones; one (FCB1) is described below. Thus, the corresponding mRNA must occur at a frequency of less than 0.1% in total NH17 poly(A)⁺ RNA and tenfold lower in wild-type MOLT-4.

Figure 1b shows FCB6 hybridizing to RNA from CD1⁺ T-cell lines (MOLT-4 and JM) homologous to stage II thymocytes and CD1⁺ T-cell hybridoma 47C7, but not to RNA from CD1⁻ lines corresponding to both earlier (RPMI-8402) and later (CCRF-HSB-2, HUT-78 and HUT-102) stages of differentiation, B lymphocytes (Daudi, Raji, WT52 and JI), other bone marrow

Fig. 1 a, SDS-polyacrylamide gel electrophoresis analysis showing the immunochemical characterization of the rabbit anti-HTA 1 α chain serum. A lysate of ¹²⁵I-surface-labelled NH17 cells was adsorbed onto the rabbit antiserum and monoclonal antibodies M241, NU-T2 and NA1/34. These monoclonal antibodies identify the three CD1 molecular subsets³⁻⁶. The immunoprecipitated antigens were treated with TFMS to cleave O-glycosyl bonds. ³⁵S-methionine-labelled products from a cell-free translation of NH17 poly(A)⁺ RNA were adsorbed onto the rabbit antiserum and as a control, onto normal rabbit serum. b, Cell-type specific expression of FCB6 sequences. Total cytoplasmic RNA from cultured cell lines was hybridized to a single-stranded ³²P-labelled probe corresponding to the 3' end of the FCB6 clone insert (downstream of the *Ava* II site at nucleotide 944). All samples gave an approximately equal signal with a *X. laevis* histone 4 probe (data not shown). c, Northern blot analysis of poly(A)⁺ RNA from MOLT-4 and HTA 1 expression mutants¹³. HLA was used as a reference of RNA loading. The HLA and FCB probes were nick-translated inserts from p001 (ref. 19) and FCB1, the FCB3' end probe as above. The positions of the 18S (1.8 kb) and 28S (5 kb) ribosomal RNAs are indicated.

Methods. NH17 cells were surface labelled by lactoperoxidase-catalyzed iodination as previously described¹³. Aliquots of a Nonidet P40 (NP40) lysate, containing up to 2×10^7 cell equivalents, were incubated for 3 h with 5 μ l of hybridoma ascites, rabbit antiserum or normal rabbit serum in a volume of 200 μ l. After addition of 100 μ l of a 1:1 suspension of Protein A-Sepharose (Pharmacia) and incubation for 30 min with shaking, immunosorbents were washed three times with PBS-0.5% NP40 and twice with 5 mM ammonium bicarbonate. All steps prior to elution were carried out at 4 °C to prevent β_2 -microglobulin dissociation⁹. Bound antigen was eluted with 50 mM diethylamine, immediately neutralized with 0.1 vol. of 0.5 M acetic acid/0.25 M ammonium bicarbonate and freeze-dried. TFMS treatment was according to Edge *et al.*¹⁴. Reactions (in 200 μ l aliquots) were allowed to proceed for 3 h at 0 °C. Cell-free translations were in a messenger-dependent rabbit reticulocyte lysate (supplied by Dr T. Hunt), containing RNA at <0.1 mg ml⁻¹ and ³⁵S-methionine (Amersham, specific activity >800 Ci mmol⁻¹) at a final concentration of 10 mCi ml⁻¹. Reactions were incubated for 2 h at 30 °C and then diluted with 9 vols. PBS-0.1% NP40 before immunoadsorption as described above. SDS-polyacrylamide gel electrophoresis was on 7.5–15% linear gradient gels. For cDNA cloning, polysomes were isolated from NH 17 cells by Mg²⁺ precipitation¹⁵. Polysomes (up to 35 O.D.₂₆₀) were immunoselected by incubation for 1 h at 4 °C with 20 μ l of an RNAase-free, Protein A-Sepharose-purified IgG fraction from the rabbit anti-HTA 1 antiserum in 250 μ l polysome buffer (250 mM NaCl, 50 mM Tris/Cl pH 7.4, 2 mM MgCl₂, 0.1% NP40, 0.1 mg ml⁻¹ heparin and 2 μ g ml⁻¹ cycloheximide). Immune complexes were bound to Protein A-Sepharose (100 μ l of a 1:1 suspension) and washed extensively in polysome buffer at 4 °C. Bound RNA was recovered by elution in 100 mM NaCl, 10 mM Tris/Cl pH 7.4, 1 mM Na₂EDTA and 0.2% SDS, phenol/chloroform extracted and ethanol precipitated. The cDNA was ligated to *Sma* I-digested, calf intestinal phosphatase treated M13mp11¹⁶. Screening by mRNA selection was as described by Miller *et al.*¹⁷ with the following modifications. 960 clones were individually grown in 96 well microtitre plates and pooled in groups of 12. Single stranded DNA was prepared from each pool and applied to a 25 mm² nitrocellulose paper square. The 80 squares were then hybridized to NH 17 poly(A)⁺ RNA, and after washing, were sorted into 10 groups, from which bound RNAs were separately eluted and ethanol precipitated. *In vitro* translation and immunoprecipitation were as described above. The human T-cell lines RPMI-8402, MOLT-4, JM, CCRF-HSB-2, HUT-78 and HUT-102, the B-cell lines Daudi and Raji, the myeloid cell line HL-60 and the non-T, non-B cell lines K-562 and U-937 are described by Minowada *et al.*^{20,21}. A 6-thioguanine resistant subline of the CCRF-CEM human T-cell line (HTA 1⁺) and the T-cell hybridoma 47C7 (HTA 1⁺) were obtained from Dr G. F. Burns. JI is a Burkitt lymphoma cell line²² and WT52 an EBV-transformed human B-cell line²³. Total cytoplasmic RNA was isolated by NP40 lysis followed by phenol/chloroform extraction²⁴. Aliquots of RNA at 1 mg ml⁻¹ were spotted onto a nylon membrane (Pall-Biodyne), and baked for 2 h at 80 °C before hybridization. An internally-labelled, single stranded probe was prepared as described by Neuberger²⁵. Hybridization was in 10% dextran sulphate, 5 $\times$ SSPE²⁶, 1% Sarkosyl for >18 h at 65 °C, washing in 2 $\times$ SSPE for 1 h at 65 °C.



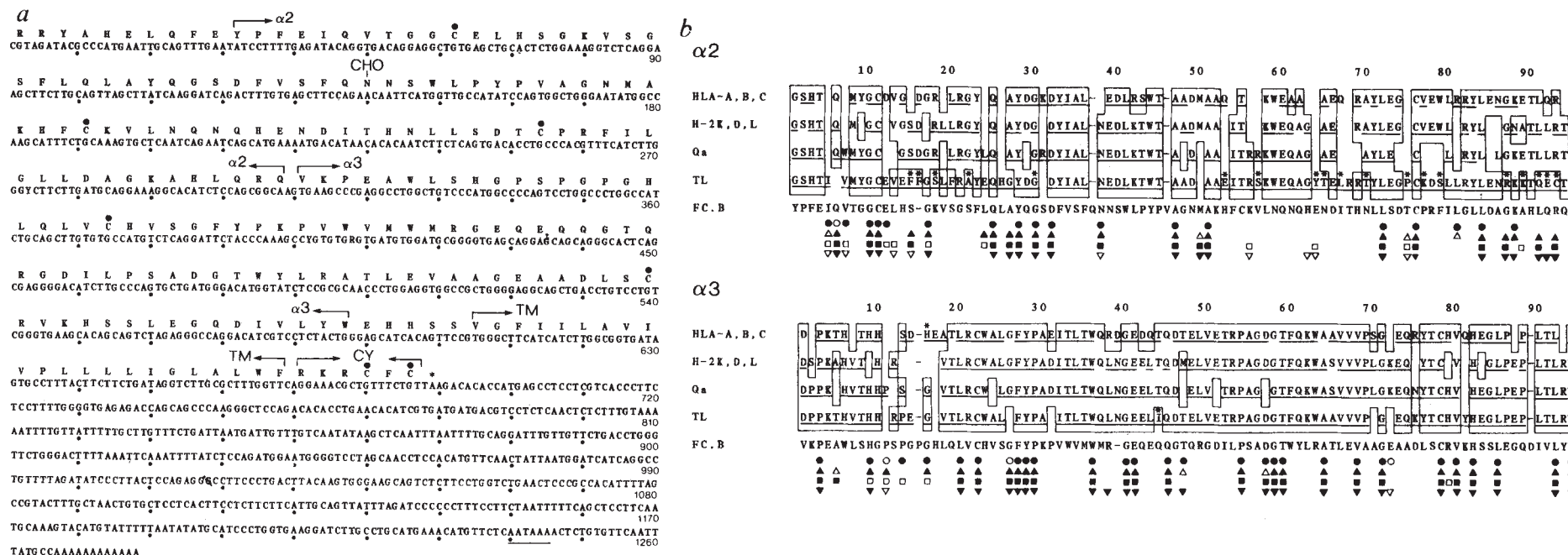


Fig. 2 *a*, Nucleotide sequence of the FCB6 and FCB1 clone inserts. The two clones are essentially identical over the whole length of FCB6, the three single base differences (T for C at positions 273 and 402, and G for A at position 408 in FCB1) probably being cloning artefacts. FCB1 is 180 nucleotides longer than FCB6 giving a total length of 1278 nucleotides, out of an estimated mRNA size of 1.4–2.1 kb (see text). In the deduced amino acid sequence the boundaries of the $\alpha 2$, $\alpha 3$, transmembrane (TM) and intracytoplasmic (CY) domains are indicated. Filled circle cysteines; CHO, single potential site for N-linked glycosylation; polyadenylation signal underlined. *b*, Alignment of the deduced FCB protein sequence with human and mouse MHC class I antigen sequences³³. These are given as consensus sequences of residues which are either invariant (underlined) or present in >80% of the individual sequences. Mouse class I genes are divided into three subclasses (H-2K,D,L, Qa and T1a) on the basis of sequence homology. Residues common to all four consensus sequences are boxed. Subclass specific residues are marked by an asterisk. Shared residues $\bullet/\circ$ = between FCB and TL; $\blacktriangle/\triangle$ = between FCB and Qa; $\blacksquare/\square$ = between FCB and H-2K,D,L; $\blacktriangledown/\triangledown$ = between FCB and HLA-A,B,C. Filled symbols, sharing with >80% of the individual sequences, open symbols with at least two. *c*, Alignment of the deduced FCB protein sequence with MHC class I α chain and class II β chain protein sequences^{33–34}. Human/rabbit (H/Rb) class I α , mouse/rat (M/Rt) class I α , human (H) class II β and mouse/rat (M/Rt) class II β are consensus sequences consisting only of residues present in >80% of the individual sequences. Residues shared across species are boxed, class-specific residues marked by an asterisk. Symbols above and below the FCB sequence indicate residues shared with the class I α and the class II β consensus respectively. Filled symbols, sharing with >80% of the individual sequences, open symbols with at least two.

Methods. Sequencing was by the shotgun method²⁷. Computer programs for handling sequencing data described in Staden²⁸. The NBRF, EMBL and Genbank databases were searched for homologies to the FCB nucleic acid and deduced protein sequences using the Wilbur-Lipman comparison program³². Whilst no significant homology exists at the nucleic acid level, at the protein level weak homology was also found to β_2 -microglobulin and to immunoglobulin C regions (κ , λ , μ , δ , α and ϵ), but to no other members of the immunoglobulin supergene family.

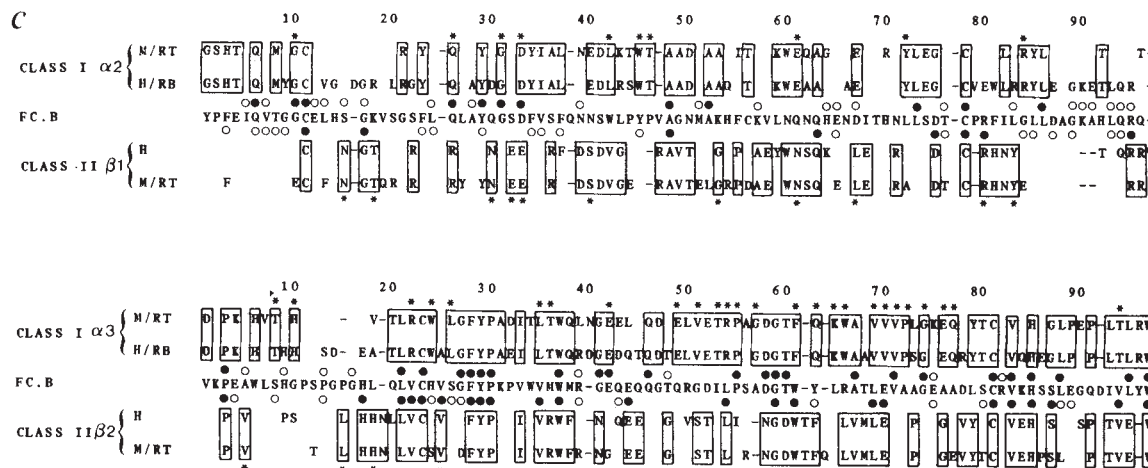


Table 1 Summary of the comparison between FCB and class I α chain amino acid sequences (a) and comparisons between FCB, class I α and class II β sequences (b)

$\alpha 2$ Domains					
	H-2				
	HLA	K,D,L	Qa	TL	
FCB	20	19	21	18	
TL	51	54	52		H-2K,D,L
Qa	57	62		75	Qa
H-2K,D,L	62		75	74	TL
		27	26	25	FCB
	TL	Qa	H-2 K,D,L	HLA	
$\alpha 3$ Domains					
	Class I α		Class II β		
	H/Rb	M/Rt	H	M/Rt	
FCB	46	37	34	38	
Class II β	M/Rt	26	22	88	
	H	24	21		
Class I α	M/Rt	107			

Data in a as in Fig. 2b. The total number of positions at which an identical residue occurs in >80% of the sequences compared is recorded. The data refer separately to the $\alpha 2$ (94 positions) and $\alpha 3$ (94 positions) domains. Data in b as in Fig. 2c. The total number of positions at which an identical residue occurs in >80% of the sequences compared is recorded. The data refer to the sum of the $\alpha 2$ and $\alpha 3$ domains (194 positions).

derived (K-562, HL-60, U-937), carcinoma (HeLa) and neuroblastoma (NB100) cells. In addition, Northern blot analysis of poly(A)⁺ RNA from NL9 (a low HTA 1 expression mutant¹³), MOLT-4 and NH17 (Fig. 1C) shows quantitative variations in FCB6-hybridizing RNAs as is observed for the HTA 1 antigen. Thus, based on the biochemical and immunological analyses and on the pattern of expression in different cell types and in MOLT-4 HTA 1 expression mutants, we conclude that FCB6 codes for a CD1 antigen.

CD1 consists of at least three distinct molecules³⁻⁶. The specificity of the rabbit antiserum we have used, as shown by immunoprecipitation, seems to rule out both the CD1 [Thy, gp45, 12] 'M241' and the CD1 [Thy, gp45, 12] 'NU-T2' types, (Fig. 1a). Thus we are left with either CD1 [Thy, gp45, 12] 'NA1/34' (HTA1/T6) or a putative fourth CD1 molecule⁶. The analysis of the MOLT-4 mutants is consistent with FCB6 encoding HTA 1, although further protein chemistry and/or gene expression studies are needed.

The nucleotide sequences of FCB6 and FCB1 are shown in Fig. 2a. The terminal poly(A) stretch is preceded 21 nucleotides upstream by a canonical polyadenylation signal (AAUAAA), while at the opposite end there is a 960 nucleotide-long open reading frame. The predicted 230 amino acid sequence accounts for 26 of the apparent 35 kDa of the HTA 1 α chain. It can be divided into domains resembling MHC class I antigen domains: a stretch of 10 amino acids, from the otherwise missing $\alpha 1$, is followed by $\alpha 2$ - and $\alpha 3$ -like domains, both 93 amino acid long,

a connecting pentapeptide, a 22-amino-acid hydrophobic (trans-membrane?) segment and a 6-amino-acid (intracytoplasmic?) tail starting with a cluster of three basic residues. The $\alpha 2$ and $\alpha 3$ domains contain a pair of cysteines separated by 64 and 54 residues respectively. In addition the $\alpha 3$ domain contains a tryptophan (residue 139 in Fig. 2a) at a position characteristic of immunoglobulin domains, and a patch of alternating hydrophobic amino acids starting 16 residues on the NH₂-side of the second-half cysteine. These features suggest a folding pattern typical of immunoglobulin type domains^{29,30}. Another cysteine is found in the $\alpha 2$ domain and two more in the intracytoplasmic tail; these could mediate the association with the T8 polypeptide^{11,12}. A single potential site for N-linked glycosylation (Asn-Asn-Ser) is found in the $\alpha 2$ domain. However five carbohydrate side chains have been postulated in MOLT-4 HTA 1³¹. If so, either the remaining four side chains are N-linked to the $\alpha 1$ domain, or some of the sugars must be bound through O-links to serine and/or threonine residues.

A computer search shows that the FCB protein is clearly related only to MHC class I α chains and class II β chains as detailed in Fig. 2b,c and Table 1. Figure 2b compares the FCB protein and human and mouse class I α chains over both the $\alpha 2$ and $\alpha 3$ domains; FCB is significantly less related to the known class I antigens than the latter are to themselves (Table 1a). In addition, FCB is not more closely related to human than to mouse class I nor is it more closely related to TL than to H-2K,D,L or Qa. Note that none of the residues which can be regarded as TL specific (marked with an asterisk in Fig. 2b) is shared with FCB. Fig. 2c compares the FCB protein, class I α chain and class II β chain 'core' sequences. FCB is almost as similar to class II β as it is to class I α . Relative to FCB, each class of MHC genes is somewhat less related to the other class, whilst across species it is much more closely related to its own class (Table 1b). Thus CD1 genes seem to have arisen before the separation of the primate and rodent branches, as the class I and class II gene ancestors diverged.

Most class I α chains are expressed in association with β_2 -microglobulin while class II β chains are not. Interestingly, whilst both the CD1 and the class II β genes have diverged from class I α genes to an almost equal extent, HTA 1 retains the association with β_2 -microglobulin, albeit weakly^{9,10}. The rabbit class I, mouse H-2D^b and possibly H-2L^d α chains do not seem to be associated with β_2 -microglobulin^{35,36}. These four alone have an arginine residue adjacent to the second-half cysteine of the $\alpha 3$ domain, a feature shared by the FCB protein. As $\alpha 3$ is the β_2 -microglobulin binding domain³⁷, this arginine may weaken the interaction with β_2 -microglobulin. This weak interaction may be important in CD1 function¹⁰.

Southern blots of MOLT-4 DNA using an FCB probe reveals multiple bands (Fig. 3a), and the same pattern is revealed by using a probe corresponding to just the $\alpha 3$ protein domain (data not shown). Thus such bands must correspond to multiple genes. This conclusion is supported by the analysis of genomic clones³⁹. Because an identical pattern is observed in 15 unrelated individuals (data not shown), these genes are likely to map to different loci, rather than being alleles. One prominent band appears for both *Eco*RI and *Hind*III digests (at 4.5 and 6.1 kbp respectively), but two appear for *Xba*I treatment (at 8.3 and 5.5 kbp); these bands should correspond to the FCB gene since an *Xba*I site is found within the FCB cDNA (at nucleotide 558). Furthermore, they are the only bands detected probing with the 3' end of FCB. The identical pattern produced by human placental and MOLT-4 DNA (Fig. 3a) suggests that CD1 expression does not require rearrangement. As mouse DNA shows only two, weaker bands, even at low stringency, only a limited conservation of CD1 sequences exists between mouse and man.

Northern blots of MOLT-4 show two main FCB-hybridizing species at 2.1 and 1.4–1.6 kb, showing parallel variations in the HTA 1 expression mutants (Fig. 1c). Both are detected with the

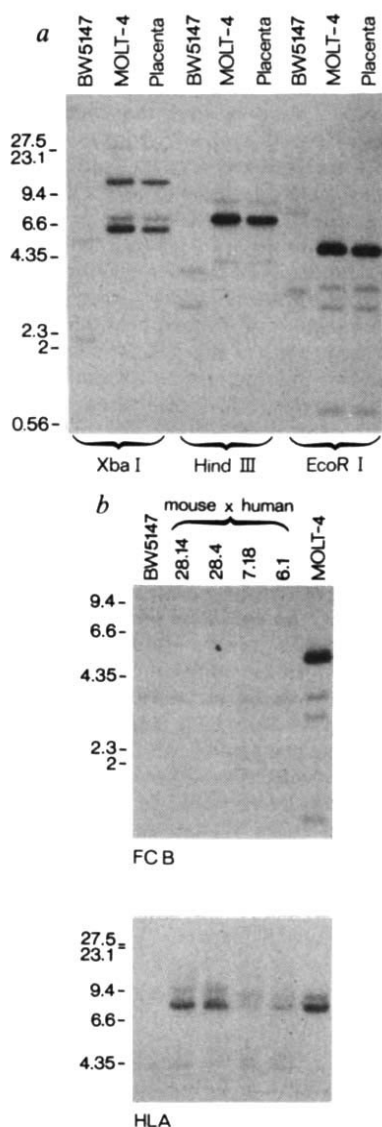


Fig. 3 *a*, Southern blot of human placental DNA, MOLT-4 and BW5147 (ref. 38) DNA probed with nick-translated whole FCB1 clone insert. The restriction enzymes used are indicated. Hybridization and washing conditions as in the legend to Fig. 1. *b*, Southern blot of EcoRI digested DNA from MOLT-4, BW5147 and four mouse $\times$ human hybrids. Clones 7-18, 28-4 and 28-14 derive from Con A-stimulated and clone 6-1 from PKW-stimulated human peripheral blood lymphocytes fused to BW5147. The probes were a nick-translated FCB1 clone insert and an internally-labelled, single-stranded probe from an M13 subclone of p001 (3' to the *Sau* 961 site at nucleotide 956). Hybridization and washing conditions, as in the legend to Fig. 1 for the FCB probe; for the HLA probe washing was in $0.1 \times$ SSPE for 1 h at 65°C . Differences in the HLA pattern between MOLT-4 and the hybrids are likely to result from restriction fragment polymorphisms.

FCB 3' end probe, which only hybridizes to one band on the Southern blot (see above). Thus the FCB gene gives rise to multiple transcripts that differ in their site of initiation and/or pattern of splicing or polyadenylation. Similar phenomena have been reported in other systems (see for example refs 40-42).

All known genes coding for β_2 -microglobulin-binding proteins cluster in the MHC. This is not true of CD1 genes. We have investigated the pattern of segregation of CD1 and MHC genes using mouse $\times$ human somatic hybrids that contain human chromosome 6 and express HLA as judged by the binding of monoclonal antibody W6/32 (data not shown). By judicious choice of probe and stringency conditions, HLA sequences are readily distinguishable on Southern blots from the cross-hybridizing mouse H-2 and are detectable in all hybrids (Fig. 3*b*, bottom panel). However, none of the FCB-hybridizing

human bands is visible (Fig. 3*b*, top panel). Thus, CD1 genes do not map to the MHC locus on chromosome 6. Preliminary experiments using other mouse $\times$ human hybrids provided by Dr J. Schroeder suggest that the CD1 genes we have described map to chromosome 1 (F. Calabi *et al.*, unpublished results), and it remains to be seen whether they are organized in a tight cluster in analogy to MHC genes.

The data presented here show that human CD1 and mouse TL antigens, although sharing some phenotypic features (pattern of cell-type expression and association with β_2 -microglobulin), are not genotypically equivalent. This is in agreement with previous data indicating that Tla genes are not conserved in evolution⁴³. However, CD1 genes appear to be relatively conserved, on the basis of the cross-hybridization between human and mouse sequences. Such mouse CD1-like genes may provide a useful model to investigate the function of this new gene family.

We thank Dr R. F. Kefford for help in raising and characterizing the rabbit anti-HTA 1 serum, J. J. Jarvis for technical assistance, Dr J. Schroeder, G. F. Burns, R. W. Knowles and K. Sagawa for cell lines and monoclonal antibodies, Drs M. Waye and N. Gay for the gift of DNAs, and Drs M. J. Sharpe, L. H. Martin, M. S. Neuberger and R. Sitia for reading the manuscript. F.C. is supported by a Fellowship from the Leukemia Society of America.

Received 17 June 1986; accepted 18 August 1986.

- Bernard, A., Boumsell, L. & Hill, C. *Joint Report of the First International Workshop on Human Leukocyte Differentiation Antigens by the Investigators of the Participating Laboratories in Leukocyte Typing* (eds A. Bernard, L. Boumsell, J. Dausset & C. Milstein) (Springer, Berlin, 1984).
- Boyse, E. A. *Cell* **38**, 1-2 (1984).
- Knowles, R. W. in *Leukocyte Typing 2* (eds Reinherz, E. L., Haynes, B. J., Bernstein, I. D. & Nadler, L. M.) 259-264 (Springer, New York, 1985).
- Olive, D., Dubreuil, P. & Mawas, C. *Immunogenetics* **20**, 253-264 (1984).
- Kahn-Perles, B., Wietzerbin, J., Caillol, D. H. & Lemonnier, F. J. *Immun.* **134**, 1759-1765 (1985).
- Amiot, M. *et al.* *J. Immun.* **136**, 1752-1758 (1986).
- McMichael, A. J. *et al.* *Eur. J. Immun.* **9**, 205-210 (1979).
- Reinherz, E. L. & Schlossman, S. F. *Cell* **19**, 821-827 (1980).
- Ziegler, A. & Milstein, C. *Nature* **279**, 243-244 (1979).
- Kefford, R. K., Calabi, F., Fearnley, I., Burrone, O. R. & Milstein, C. *Nature* **308**, 641-642 (1984).
- Ledbetter, J. A., Tsu, T. T. & Clark, E. A. *J. Immun.* **134**, 4250-4254 (1985).
- Snow, P. M., van de Rijn, M. & Terhorst, C. *Eur. J. Immun.* **15**, 529-532 (1985).
- Burrone, O. R., Calabi, F., Kefford, R. F. & Milstein, C. *EMBO J.* **2**, 1591-1595 (1983).
- Edge, A. S. B., Faltynek, C. R., Hof, L., Reichert, L. E. Jr & Weber, P. *Analyt. Biochem.* **118**, 131-137 (1981).
- Palmiter, R. D. *Biochemistry* **13**, 3606-3615 (1974).
- Messing, J. *Meth. Enzym.* **101**, 20-78 (1983).
- Miller, J. S., Paterson, B. M., Ricciardi, R. P., Cohen, L. & Roberts, B. E. *Meth. Enzym.* **101**, 650-674 (1983).
- Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA cloning I* (ed. Glover, D. M.) 49-78 (Cambridge University Press, 1985).
- Sood, A. K., Pereira, D. & Weissman, S. *Proc. natn. Acad. Sci. U.S.A.* **78**, 616-620 (1981).
- Minowada, J. *J. Cancer Res. clin. Oncol.* **101**, 91 (1980).
- Minowada, J., Sagawa, K., Trowbridge, I. S., Kung, P. D. & Goldstein, G. in *Malignant Lymphomas: Etiology, Immunology, Pathology and Treatment* (eds Kaplan, H. & Rosenberg, S.) 53-74 (Academic, New York, 1982).
- Bernheim, A., Berger, R. & Lenoir, G. *Cancer Genet. Cytogenet.* **3**, 307-315 (1981).
- Trucco, M. M., Garotta, G., Stocker, J. W. & Ceppellini, R. *Immun. Rev.* **47**, 219-252 (1979).
- Calabi, F. & Neuberger, M. S. *EMBO J.* **4**, 667-674 (1985).
- Neuberger, M. S. *EMBO J.* **2**, 1373-1378 (1983).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular cloning. A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Bankier, A. T., Weston, K. M. & Barrell, B. G. *Meth. Enzym.* (in the press).
- Staden, R. *Nucleic Acids Res.* **14**, 217-231 (1986).
- Svasti, J. & Milstein, C. *Biochem. J.* **126**, 837-850 (1972).
- Williams, A. F., Barclay, A. N., Clark, M. J. & Gagnon, J. in *Gene Expression During Normal and Malignant Differentiation* (eds Andersson, L. C., Gahrberg, C. G. & Ekblom, P.) 125 (Academic, London, 1985).
- van Agthoven, A. & Terhorst, C. *J. Immun.* **128**, 426-432 (1982).
- Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726-730 (1982).
- Klein, J. & Figueroa, F. *Immun. Today* **7**, 41-44 (1985).
- Klein, J. & Figueroa, F. *Immun. Today* **7**, 78-81 (1985).
- Wilkinson, J. M., Tycocinski, M. L., Coligan, J. E., Kimball, E. S. & Kindt, T. J. *Molec. Immun.* **19**, 1441-1451 (1982).
- Potter, T. A., Boyer, C., Schmitt Verhulst, A., Golstein, P. & Rajan, T. V. *J. exp. Med.* **160**, 317-322 (1984).
- Yokoyama, K. & Nathanson, S. G. *J. Immun.* **130**, 1419-1425 (1983).
- Hyman, R. & Stallings, V. *Immunogenetics* **6**, 447-458 (1978).
- Martin, L. H., Calabi, F. & Milstein, C. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Bernard, O., Cory, S., Gerondakis, S., Webb, E. & Adams, J. M. *EMBO J.* **2**, 2375-2383 (1983).
- Parnes, J. R., Robinson, R. R. & Seidman, J. G. *Nature* **302**, 449-452 (1983).
- Blattner, F. R. & Tucker, P. W. *Nature* **307**, 417-422 (1984).
- Rogers, J. H. *EMBO J.* **4**, 749-753 (1985).

Cell-type specific protein binding to the enhancer of simian virus 40 in nuclear extracts

Irwin Davidson, Catherine Fromental, Patrick Augereau, Alan Wildeman*, Martin Zenke* & Pierre Chambon†

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, Strasbourg, France

Enhancers are *cis*-acting activators of transcription from homologous or heterologous promoter elements of viral and cellular genes (see refs 1–6 for reviews). The activity of the simian virus 40 (SV40) (refs 7–9) and immunoglobulin heavy-chain gene (IgH) (refs 10, 11) enhancers has been reproduced to some extent *in vitro* and appears to be mediated by *trans*-acting factors both *in vitro*^{7–11} and *in vivo*^{12,13}. The SV40 enhancer consists of multiple sequence motifs in two domains, A and B (Fig. 1, see ref. 14): domain B contains GT-I and -II and two TC motifs, of which only TC-II is important for enhancer activity in HeLa cells^{14,15}; and domain A contains the P and the two Sph motifs, the repetition of which generates the sequence 5'-ATGCAAAG-3', similar to the 'octameric' sequence of the IgH enhancer, (Fig. 4i; refs 14, 16), where it is important for enhancing activity^{17,18}. Each SV40 enhancer motif is a binding site for a protein or proteins present in HeLa cell nuclear extracts^{19,20}. Unlike the SV40 enhancer, which is active in HeLa and lymphoid B cells, the IgH enhancer is preferentially active in B cells^{13,17,18,21,22}, suggesting that not all

the *trans*-acting factors necessary for its activity are present in HeLa cells. However, the IgH enhancer can compete with the SV40 enhancer *in vitro* in HeLa or lymphoid cell extracts^{10,23} and *in vivo* in B cells¹³. Here we show that both human HeLa and BJA-B²⁴ lymphoid B-cell nuclear extracts contain proteins that bind to specific, sometimes overlapping, motifs of the SV40 enhancer. Some binding is cell-specific, suggesting that it is not the same set of sequence motifs and proteins that is responsible for the enhancer activity in the two cell types. This is confirmed by our results obtained *in vivo* with mutated SV40 enhancers.

In DNase I footprinting assays, both HeLa and BJA-B extracts generate regions of protection and hypersensitivity over the SV40 enhancer (pAO in Fig. 2A, B; summarized in Fig. 1). In HeLa extracts this pattern is essentially as previously described^{19,20}, whereas it is significantly different in BJA-B extracts (in both cases the 21-base pairs (bp) repeat region was similarly protected by the binding of factor Sp1^{25,26}). The exact locations of the DNase I hypersensitive sites at the border between domains A and B differed using BJA-B (LE3, LL4) and HeLa (E3, L5) extracts. As with HeLa extracts, protection was observed in BJA-B extracts over the P motif (LL7, LE6) and the Sph-octameric motifs (LL5, LE4), but the exact boundaries of the protected regions are not the same as in HeLa extracts (L6, E4 and E6), and an additional hypersensitive site (LL6) was observed. In domain B, motifs TC-II and GT-I (LL3, LL2 and LE2) were also protected using BJA-B extracts. The HeLa cell hypersensitive site L3 is much less evident with BJA-B extracts (Fig. 2B, dashed arrow), which induced two new hypersensitive sites (positions 242 and 244, Fig. 2B filled arrows) and the protected region (L2 versus LL2) is larger in HeLa extracts. A third protected region is seen in BJA-B extracts on both strands, partially overlapping with the protection seen in HeLa extracts over motif GT-II (L1), but extending further upstream (LL1, LE0 in Fig. 2A, B; see Fig. 1 and data not shown), and on the early strand a BJA-B specific hypersensitive site LE1 was observed.

* Present addresses: Department of Molecular Biology and Genetics, University of Guelph, Ontario, Canada N1G 2W1 (A.W.). European Laboratory of Molecular Biology, Postfach 102240, 6900 Heidelberg 1, FRG (M.Z.).

† To whom correspondence should be addressed.

Fig. 1 Schematic representation of the result of DNase I footprinting and DMS methylation protection experiments over the SV40 enhancer in HeLa and lymphoid cell nuclear extracts. Top, the SV40 early promoter region containing a single copy of the 72-bp sequence from plasmid pAO together with some relevant restriction sites. The early and late messenger RNA coding strands of the 72-bp and 5' flanking sequences are presented using the BBB nucleotide numbering system^{14,29}. Regions of protection from DNase I digestion in HeLa and lymphoid cell nuclear extracts are shown by solid black lines above and below the sequence of each strand, respectively. Broken lines flanking regions E2 and E4, residues which, although not clearly protected in pAO, are deprotected by point mutations in GT-I and Sph-II, respectively, and thus appear to be part of these regions²⁰. Sites rendered hypersensitive to DNase I digestion by each extract are indicated by the large (strong) and small (weaker) arrows (dotted arrow, see text). The location of the G residues which were protected from methylation by DMS in the HeLa (open triangles) and BJA-B (open circles) cell extracts are shown with the hypermethylated G residues (filled triangles and circles for HeLa and BJA-B extracts, respectively). Note that, using the HeLa cell extract, efficient protection of G189 within the P motif of pAO, occurs only when methylation is performed at 0 °C (data not shown; Fig. 3 legend). The nomenclature of each region is that used in Fig. 2A, B. Locations of enhancer sequence motifs essential for *in vivo* activity in HeLa cells¹⁴ (GT-I, GT-II, TC-II, Sph-I, Sph-II and P; GT-I corresponds to the enhancer core sequence described by Weiher *et al.*¹⁵) are shown with those of the other enhancer motifs (TC-I and octamer) and of domains A and B. GT-I, GT-II, Sph-I, Sph-II and P are boxed with solid lines and the octameric motif with broken lines. The locations, and DNA sequence (late strand) of the mutations present in pA20–28 mutants of pAO¹⁴ are shown between the early and late mRNA coding strands. Only those DNase I and DMS 'footprints' which were consistently observed in several experiments using different preparations of HeLa or BJA-B cell extracts are indicated.

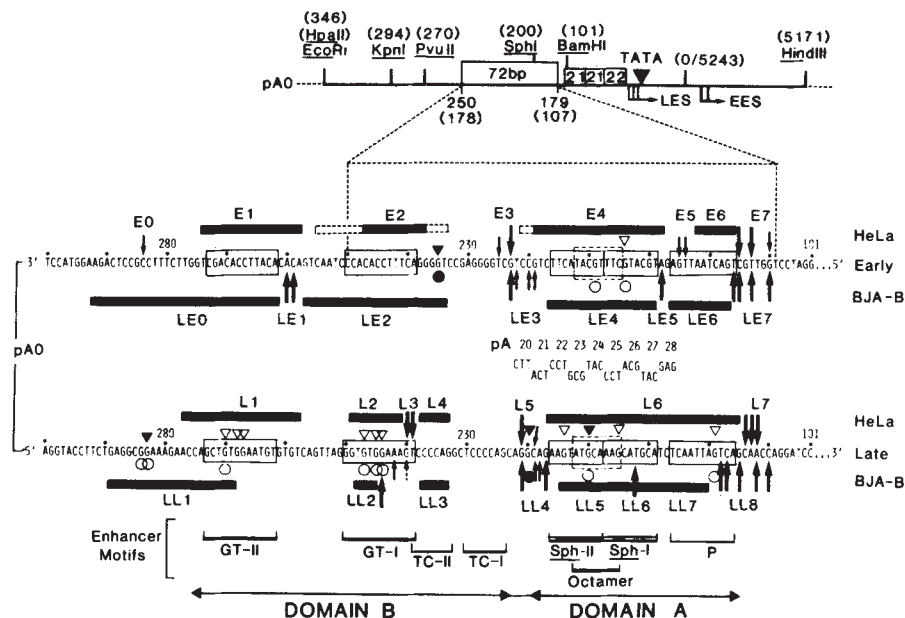
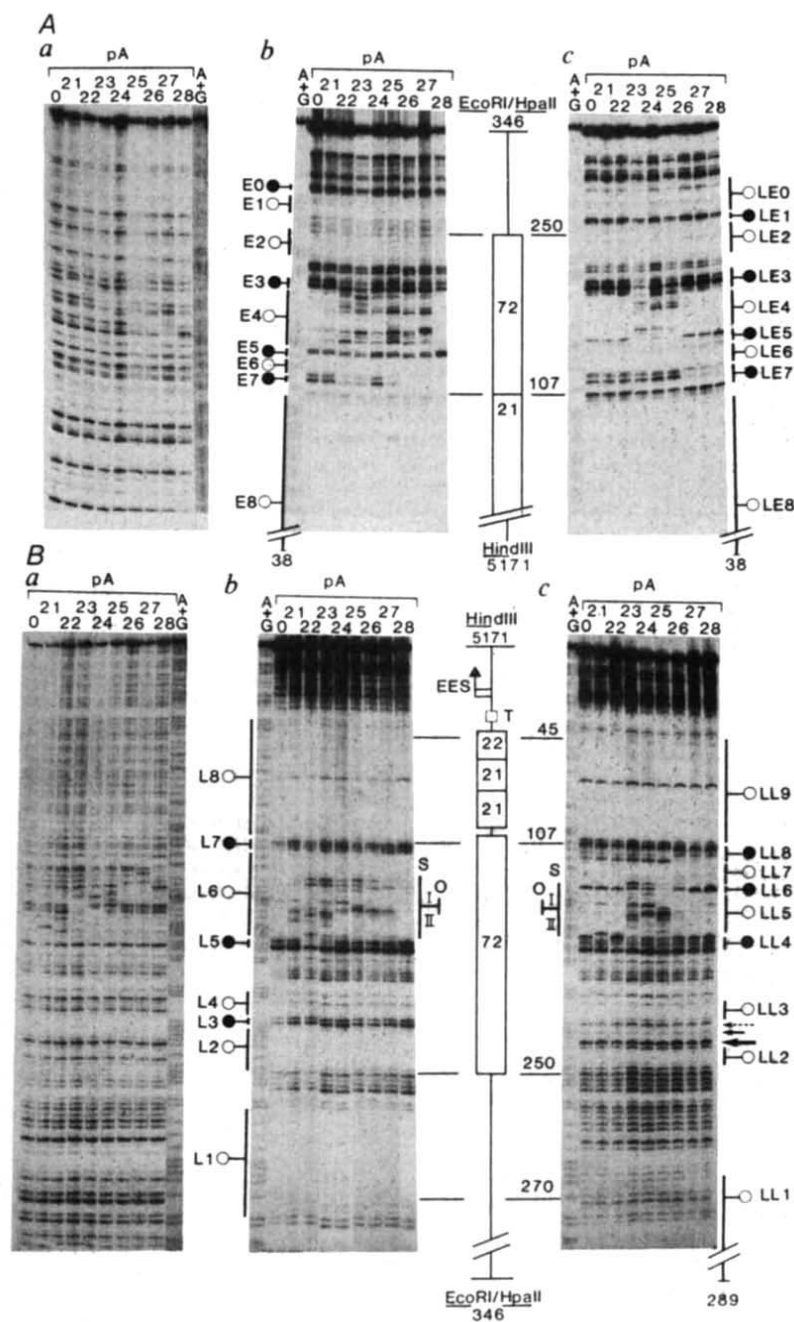


Fig. 2 DNase I footprinting over the SV40 enhancer of pAO and pA21-pA28 in HeLa and BJA-B cell nuclear extracts. **A**, Early strand. **a**, Naked DNA; **b**, HeLa extract 6 μ l; **c**, BJA-B extract 6 μ l. **B**, Late strand. **a**, Naked DNA; **b**, HeLa extract 6 μ l; **c**, BJA-B extract 6 μ l. **A**, The 21-bp repeat region (21) and the unique 72-bp (72) sequence are indicated along with the positions of the *Hind*III and *Eco*RI sites (see Fig. 1). The regions whose sensitivity to DNase I digestion was increased (solid circles) or decreased (open circles) by the HeLa and BJA-B cell nuclear extracts are indicated on the left of the central panel and to the right of the figure, respectively. E, early strand; LE, lymphoid early strand. In each case the coordinates of the downstream boundary of the region LEO/LE8 are indicated (38). As the upstream boundary of the region LEO is at the limit of the resolution of the gel the early strand is also labelled at the engineered *Sal*I site between the 21-bp repeat region and the TATA box in plasmid pSEGO (ref. 30). DNase I footprinting experiments using this fragment allow the position of this boundary to be determined (data not shown). **B**, Symbols as in **A**. L, late strand; LL, lymphoid late strand. The positions of the unique 72-bp sequence (72) the 21-bp repeat region (21, 21, 22), the TATA-box (T) and the early-early RNA start sites (EES) are indicated (see Fig. 1). S-I and -II, Sph-I and -II; O, octameric motif. The lymphoid extract-specific hypersensitive sites in GT-I (see text) are indicated by the large (position 244) and small (position 242) arrows to the left of the figure. The coordinate of the upstream boundary of region LL1, determined in similar experiments where electrophoresis was performed for a shorter time (data not shown), is also indicated. Because the lanes corresponding to pAO, pA22 and pA25 are noticeably less intense in the HeLa extract (compare the intensities at the top of the gel or at L7), corrections are made when comparing the effect of the mutations. Thus the deprotection at L6 in pA22 is equivalent to that in pA23 and the deprotection of Sph-I in pA25 is equivalent to that in pA22 and pA23. In **A** and **B** the minor variations in the positions of the DNase I cleavage sites in each mutant can be ascribed to the changes in the pattern of digestion of the corresponding naked DNA. In **A** and **B** only those DNase I footprints consistently observed in several experiments using different preparations of extracts are indicated.

Methods. Nuclear extracts from HeLa and BJA-B cells were prepared as previously described^{8,20} and their protein concentration varied within a twofold range. The early (**A**) and late (**B**) mRNA coding strands were labelled by first digesting the appropriate recombinant plasmid (pAO; pA21-28) with *Hind*III and *Eco*RI, respectively, followed by phosphorylation of these sites using T4 polynucleotide kinase and [γ -³²P]ATP (5,000 Ci mmol⁻¹). After recleavage with *Eco*RI and *Hind*III, respectively, the fragments containing the uniquely end-labelled early promoter region were electroeluted from 5% polyacrylamide gels. DNase I footprinting reactions were performed essentially as described by Wildeman *et al.*²⁰. Briefly, the nuclear extract (indicated in the figure) was preincubated on ice for 15 min with 25 ng linearized pBR322 DNA in a volume of 10 μ l containing 20 mM HEPES pH 7.9, 30 mM KCl, 3 mM MgCl₂ and 10% glycerol. For the naked DNA the nuclear extract was replaced by an equal volume of nuclear extract dialysis buffer. Following the preincubation 10–20 $\times 10^3$ c.p.m. (~5 fmol) of the labelled DNA was added and incubation continued at 20 °C for 10 min. Freshly diluted precalibrated DNase I (1–2 μ l) was added and the digestion allowed to continue for 90 s. The reaction was terminated by the addition of 200 μ l 0.3% SDS, 0.13 M NaCl and the samples deproteinized by extraction with phenol and chloroform. The digestion products were ethanol-precipitated and separated on a 8% denaturing polyacrylamide gel. An A+G sequence ladder of the pAO template was electrophoresed in parallel.



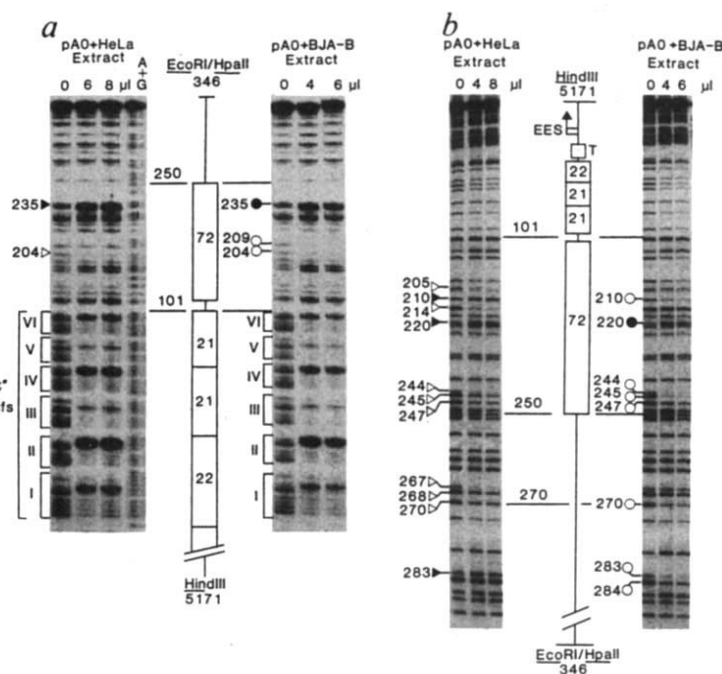
in vivo enhancing activity in HeLa cells¹⁴. But the pattern of close contact sites is different in BJA-B extracts, for example, the overlapping Sph and octameric motifs of domain A (see pAO in Figs 1, 4i). In HeLa extracts a protein (or proteins) appears to bind to Sph-I and -II (close contact sites G205 and G214) but not to the octameric motif (no contact sites at G209 and G210). In contrast, a protein (or proteins) in BJA-B extracts appears to bind only to the octameric motif (contact sites G209 and G210 lie in this motif and G204 is located immediately adjacent). Because of the complexity of the region containing the overlapping Sph and octameric motifs, mutated pAO templates (pA21-28, see Figs 1, 4i; ref. 14) were used to investigate further protein binding in this region.

Mutations in Sph-II (pA22-24) cause a deprotection of Sph-I and -II on both strands in HeLa extracts (Fig. 2A, B; refs 19, 20). Mutations in Sph-I (pA25-27) deprotect both Sph motifs on the early strand but only Sph-I on the late strand. In contrast

Methylation protection experiments²⁷ were performed to determine which G residues are protected by close contact with the proteins (Fig. 3a, b). In HeLa extracts G205 and G214, which lie at identical positions in motifs Sph-I and -II, are protected, whereas G210 and G220 are hypermethylated. In contrast, G210 in the overlap region between the Sph-II and octameric motifs is protected in BJA-B extracts, although G220 is also hypermethylated. On the early strand G204 and G209 are protected in BJA-B extracts, whereas only G204 is protected in HeLa extracts and G235 within TC-II is strongly methylated in each extract. Both extracts protect G244, G245 and G247 on the late strand in GT-I (Fig. 3b), but protection of G245 is more efficient in BJA-B extracts. In GT-II, HeLa extracts protect G267, G268 and G270, whereas G270, G283 and G284, but not G267 or G268, are protected in BJA-B extracts. These experiments (summarized in Fig. 1) demonstrate that proteins make specific close contacts with all except one (TC-II) of the motifs essential for

Fig. 3 DMS methylation protection over the SV40 enhancer of pAO in HeLa and BJA-B cell nuclear extracts. *a*, Early strand; *b*, late strand. *a*, The quantity and origin of the nuclear extracts used are indicated at the top. The locations of the protected G residues within the enhancer region are indicated with open triangles (HeLa extract) or open circles (BJA-B extract). Hypermethylation of G235 in the cluster of 4 G residues located at positions 235–238 is indicated using a filled triangle and a filled circle for HeLa and BJA-B cell extracts, respectively. The location of this hypermethylation was confirmed in similar experiments using the *EcoRI/SalI* fragment from pSEGO (ref. 30), as described in Fig. 2. The positions of the unique 72-bp sequence (72) and the 21-bp repeat region (21, 21, 22) are indicated between the two panels. The coordinates of the protected and hypermethylated G residues of the GC-rich motifs (I–VI) in the 21-bp region are essentially identical to those previously described (see text). Note that GC-I is the least efficiently protected, in agreement with previous reports^{25,26,30}. *b*, Symbols as in *a*. When comparing lanes 0, 4 and 6 (μ l) using the BJA-B extract, note that all of the G residues in the third lane (6 μ l) are less intense than those in the first and second lanes (0 and 4 μ l). In both *a* and *b*, only those G residues whose methylation pattern was reproducibly modified (compared to that of neighbouring G residues) using several preparations of nuclear extract, are indicated (see Fig. 1 legend). Similar results were observed when the methylation reaction was performed at 0 °C, with the exception of G189 in the P motif which, using the HeLa extract, was efficiently protected only at 0 °C (data now shown).

Methods. The [³²P]5'-end labelled SV40 early promoter fragments were prepared as described in Fig. 2 legend. Methylation protection experiments were performed as follows: after preincubation of the nuclear extract on ice and addition of the labelled DNA as described in Fig. 2 the samples were incubated at 20 °C for 10 min; 2 μ l DMS was then added and methylation allowed to proceed for 1 min; the reaction was terminated by the addition of 200 μ l of a solution containing 1.5 M β -mercaptoethanol, 0.3 M sodium acetate and 50 μ g ml⁻¹ yeast RNA. Samples were immediately extracted with phenol/chloroform and ethanol precipitated. The methylated DNA was reprecipitated a second time and cleaved at the position of the modified G residues by treatment with piperidine at 90 °C for 30 min as described³¹. After lyophilization overnight the samples were resuspended in 200 μ l 0.3 M sodium acetate and precipitated with ethanol. The reaction products were analysed on 8% denaturing polyacrylamide gels. An A+G sequence ladder of pAO was electrophoresed in parallel.



in BJA-B extracts, deprotection is observed only when the mutations lie in the octameric motif (pA23, 24 and 25), confirming that HeLa extracts contain a protein(s) that binds to the Sph motifs, whereas the BJA-B extracts contain a protein(s) that binds to the octameric motif. No protein in the BJA-B extracts binds to the Sph motifs, as pA24 is completely deprotected whereas it is only partially deprotected in HeLa cell extracts (compare pA23 and 25 with pA24 on both strands; see also ref. 20). On the other hand, no proteins apparently bind to the SV40 octameric motif in HeLa extracts as no protection is observed using the mutants where binding to the Sph motifs is weakened (pA21, 26 and 27) or completely prevented (pA22), although under identical conditions the octameric motif of the same mutant is fully protected by BJA-B extracts. These conclusions are fully supported by the results of dimethyl sulphate (DMS) methylation-protection experiments using the same set of mutated templates (Fig. 4h and data not shown; see Fig. 4i). Note that in HeLa extracts, mutations in Sph-II deprotected the G residues in Sph-I and -II (pA22 and pA23, Fig. 4h and data not shown; see Fig. 4i), whereas partial protection of G214 in Sph-II occurs even when Sph-I is mutated (see pA26 in Fig. 4i). Similarly in pA24, which contains a mutation at the border of Sph-II, G214 is partially protected in HeLa extracts, whereas G205 of Sph-I is completely deprotected (neither of these pA24 G residues is protected by BJA-B cell extracts, data not shown). It is not known whether these and the DNase I results described above reflect the cooperative binding of two, possibly different, proteins to Sph-II and -I or the binding of a single protein which has a higher affinity for Sph-II than for Sph-I.

The recent finding that HeLa extracts contain a protein (or proteins) that binds to the SV40-like octameric motifs of the IgH enhancer (see Fig. 4i) and of the kappa light chain promoter

upstream element²⁸ suggests that HeLa extracts contain some component(s) preventing the binding of such a protein to the SV40 octameric motif. As expected, a DNase I protected region centred over the octameric motif (positions 541–548) occurs on both strands of the IgH enhancer in either HeLa or BJA-B extracts (Fig. 4a–d). The strong methylation protection of G545 in either extract and the weak protection of G546 and 551 confirms that the binding occurs on the octameric motif (Fig. 4e,f; see also Fig. 4i). Interestingly, G545 is the residue protected *in vivo* in lymphoid B cells¹⁶. We performed DNase I footprinting using a mixture of the BJA-B and HeLa extracts and pAO or pA22. With increasing amounts of BJA-B extracts and decreasing amounts of HeLa extracts, there is a progressive change in the DNase I footprint of pAO from that seen in the HeLa extracts to that characteristic of BJA-B extracts (Fig. 4g, note the appearance of the BJA-B-specific cleavage site LL6, the change from HeLa L5 to BJA-B LL4, the decrease in HeLa-specific cleavage site L3 and the increase of BJA-B-specific cleavage site at position 244 (solid arrow), all of which are already seen in lanes 2–4). The quantity of octamer-binding factor(s) present in 1 μ l of BJA-B extracts efficiently protects the SV40 octameric motif of mutant pA22, even in the presence of an excess of HeLa extract (6 μ l), whereas under similar conditions no protection occurs with HeLa extract alone (Fig. 4g, lanes 9, 10). Thus, HeLa extracts do not appear to contain an inhibitor preventing the BJA-B octamer-binding protein from interacting with the SV40 motif. Note that addition of 1 μ l BJA-B extracts to pA22 results in the full appearance of the characteristic BJA-B DNase I pattern only in the Sph-octameric motif region (compare lanes 9, 10, 12 in Fig. 4g), whereas in the other motifs of pA22, and in all the motifs of pAO, greater amounts of BJA-B extracts are required to modify

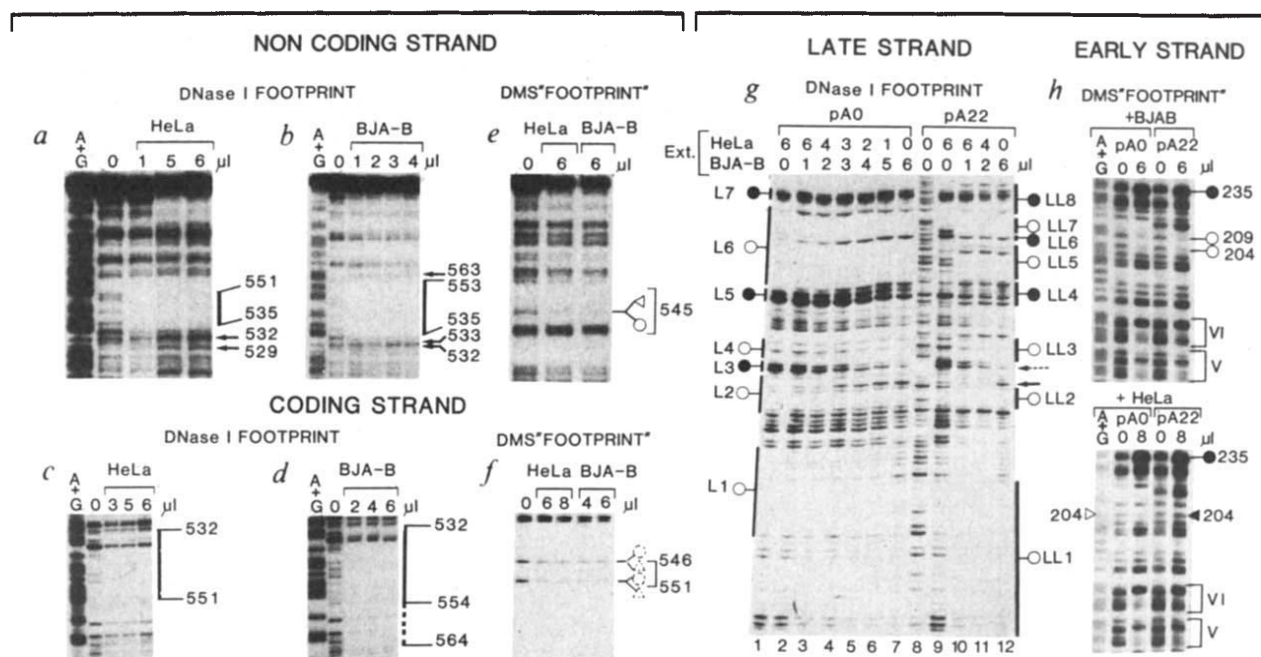
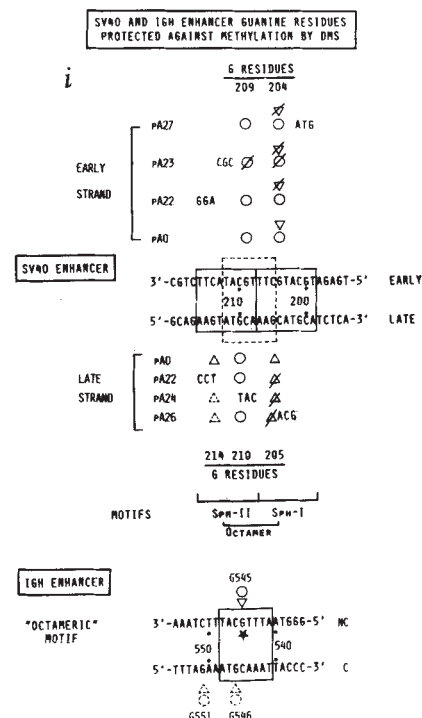


Fig. 4 *a-f*, DNase I footprinting and DMS methylation protection on the octameric motif of the IgH enhancer; *g*, DNase I footprinting on the SV40 enhancer of pAO and pA22 using either HeLa or BJA-B cell extracts or a mixture of the two; *h*, DMS methylation over the Sph-octameric region of the enhancers of pAO and pA22; *i*, comparison of the SV40 and IgH enhancer guanine residues in the Sph-octameric motifs protected against methylation by DMS. *a-f*, The preparation of the [32 P]5'-end-labelled mRNA coding and non-coding strands of the *Pst*I/*Ava*II fragment from the mouse IgH enhancer (coordinates 376-586, using the numbering system of Ephrussi *et al.*¹⁶, was performed as previously described¹¹. DNase I footprinting (*a-d*) and DMS methylation protection experiments (*e,f*) were performed as described in Figs 2 and 3 legends, respectively. In each case the origin and quantity of the nuclear extract used is indicated at the top of the figure. *a-d*, Regions protected from DNase I digestion indicated by the vertical lines to the left; coordinates according to Ephrussi *et al.*¹⁶. Arrows, sites hypersensitive to DNase I digestion. *e,f*, Symbols as described in Fig. 3 legend; broken circles and triangles indicate partial protection. *g*, The SV40 early promoter region [32 P]5'-end labelled on the late strand was prepared as described in Fig. 2 and DNase I footprinting experiments performed using a mixture of HeLa and BJA-B cell extracts. The DNA template and the quantity of each nuclear extract used are indicated at the top. The HeLa and BJA-B cell-specific regions of DNase I protection and hypersensitivity are shown in the left- and right-hand sides, respectively, using the same nomenclature as in Fig. 2. The band present in the 3' region of LL5 in lanes 10-12, of much lower intensity in other experiments (see Fig. 2b, lanes pAO-22), is not indicative of deprotection of the octamer motif as shown by the DMS methylation protection experiments (see below). *h*, DMS methylation-protection experiments on the SV40 early promoter region [32 P]5'-end-labelled on the early strand were performed as described in Fig. 3a legend. Symbols, as Fig. 3a except the closed triangle in the lower panel showing the position of G204 which in pA22 is no longer protected from methylation in HeLa cell extracts. *i*, Sequence of SV40 early and late mRNA coding strands in the region of the overlapping Sph and octameric motifs (boxed with solid and broken lines, respectively). For each mutant the position and sequence of the mutation is indicated (see also Fig. 1, ref. 14). To facilitate the comparison with the SV40 enhancer, the IgH enhancer (numbering system as in ref. 16) is orientated so that the octameric motifs of the two enhancers are directly aligned. The G residues efficiently protected by the HeLa and lymphoid cell nuclear extracts are indicated using the same symbols (open triangles and circles, respectively) as in Figs 1, 3. Symbols for partial protection by the HeLa and lymphoid cell extracts (broken triangle and circle, respectively) are the same as *a-h*. Deprotection of G residues by the mutations is indicated by the presence of a line barring the symbol. Using each extract no novel protected G residues (a protected G residue not already observed in pAO) were observed in any of the mutant templates. Star, location of the G residue of the IgH octameric motif protected from methylation *in vivo* in B cells¹⁶.

the HeLa pattern of protection (lanes 1-7, 12). Thus the SV40 octameric motif is readily available for protein binding in mutant pA22, whereas some competition may occur between the proteins that bind specifically to the Sph and octameric motifs in the wild-type pAO.

In summary, several sequence motifs of the SV40 enhancer are cell-specifically recognized using nuclear extracts. This is particularly clear for proteins that bind to domain A, as there is no evidence for protein binding to the Sph motifs in BJA-B extracts, whereas there is apparently no protein that can recognize the SV40 octameric motif in HeLa extracts. These conclusions are confirmed by *in vivo* experiments that demonstrate that the Sph motifs, but not the octameric motif, are functional in HeLa cells, and that the converse situation exists in lymphoid B cells (Fig. 5a-d). All mutations located in Sph-I and -II are detrimental to enhancer activity in HeLa cells except pA24,

which lies within the region of overlap between the Sph-II and octameric motifs. In contrast, only mutations located within the octameric motif (pA23, 24, 25) reduce the activity of the enhancer in B cells. However, both HeLa and BJA-B extracts contain a protein or proteins that can interact with the IgH enhancer octameric motif (see above), which suggests strongly that, although the 'core' octameric motif is conserved in the SV40 and IgH enhancers, the sequences recognized in each case are different, pointing to the importance of surrounding sequences (note that the protected IgH HG551 residue has no counterpart in the SV40 octameric motif and vice versa for the SV40 G204 and G205 residues, see Fig. 4i). Because DNase I footprinting titration experiments (not shown) indicate that BJA-B extracts are only threefold richer than HeLa extracts in IgH octamer-binding protein, it is unlikely that a lower affinity of this protein for the SV40 octameric motif could account for the present



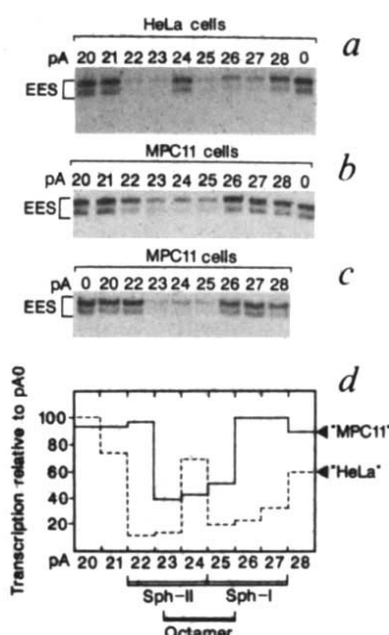


Fig. 5 Comparison of the effect of mutations pA20-28 on the activity of the SV40 enhancer in HeLa and lymphoid B cells. Transient expression assays and quantitative S_1 nuclease mapping were performed as described previously¹⁴ using calcium phosphate transfection and the recombinant p β (244+) β as an internal control. *a-c*, EES is the single-stranded probe¹⁴ fragments of 130-137 nucleotides in length protected from S_1 nuclease digestion by RNA initiated at the early-start sites of the SV40 early promoter following transfection into HeLa cells (*a*) or MPC11 BU4 mouse plasmacytoma B cells (*b, c*). In each case the transfected plasmid (pAO or one of the pA20-28 mutants, see Fig. 1) is indicated at the top of the lane. *b, c*, Results from two independent transfections with two plasmid preparations. *d*, Activity of the enhancer of mutants pA20-28 relative to the wild-type pAO. Values shown for the relative transcriptional activity of the mutants pA20-28 represent the average values of 3 (HeLa) or 4 (MPC11 BU4) independent transfection experiments with different plasmid preparations. These values were obtained by densitometric scanning of autoradiograms (exposed for various times) similar to those shown in *a-c*, followed by normalization to transcription (not shown) from the co-transfected control p β (244+) β plasmid (see ref. 14). The relative activity of each mutant is expressed as a percentage of the activity of pAO in the corresponding cell type. Solid and broken lines, relative activity of the enhancer in MPC11 BU4 and HeLa cells, respectively. Locations of Sph-I and -II and the octameric motif relative to the mutations are indicated at the bottom (see also Fig. 1).

results, which are more readily explained by the presence in HeLa and BJA-B cells of proteins exhibiting different octamer-binding properties. Post-translational modifications and/or interactions with other cell-specific proteins, leading to alterations in the specificity of binding of a single octamer protein would, however, also explain our observations (the same remark applies to the apparent absence of the Sph motif-binding protein in lymphoid cells, see above).

Proteins may also bind in a cell-specific manner to different, overlapping motifs in domain B of the SV40 enhancer (see Fig. 1 for summary). In particular, motif GT-II is only partially protected in BJA-B extracts, but there is an overlapping sequence located further upstream in close contact with a protein that specifically binds in BJA-B extracts.

Obviously, the combinatorial possibilities of regulating the activity of a given enhancer are enormously increased if the multiple sequence motifs that constitute this enhancer can interact with proteins in a cell-specific manner. These possibilities are further increased if some of these proteins, which may be

present in various ratios in different cell types, compete for binding to overlapping motifs similar to the SV40 Sph-octameric motifs described here.

We thank C. Werlé and B. Boulay for the illustrations; our secretarial staff for typing the manuscript; and the cell culture group for technical assistance. Supported by CNRS (ATP 6984), INSERM, the Fondation pour la Recherche Médicale, the Ministère de la Recherche et de la Technologie (84V0803) and the Association pour le Développement de la Recherche sur le Cancer. I.D., A.W. and M.Z. were recipients of fellowships from the Royal Society, the Natural Sciences and Engineering Research Council of Canada and the CNRS/Deutsche Forschungsgemeinschaft, respectively.

Received 17 March; accepted 11 August 1986.

1. Khoury, G. & Gruss, P. *Cell* **33**, 313-314 (1983).
2. Yaniv, M. *Biol. Cell* **50**, 203-216 (1984).
3. Chambon, P. *et al.* in *Recent Progress in Hormone Research* **40**, 1-42 (1984).
4. Serfling, E., Jasin, M. & Schaffner, W. *Trends Genet.* **1**, 224-230 (1985).
5. Wasylyk, B. in *Maximizing Gene Expression* (eds Reznikoff, W. & Gold, L.) 79-99 (Butterworth, London, 1986).
6. Picard, D. *Oxford Surv. Eukaryotic Genes* **2**, 24-48 (1985).
7. Sassone-Corsi, P., Dougherty, J., Wasylyk, B. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 308-312 (1984).
8. Wildeman, A. G., Sassone-Corsi, P., Grundström, T., Zenke, M. & Chambon, P. *EMBO J.* **3**, 3129-3133 (1984).
9. Sergeant, A., Bohmann, D., Zentgraf, H., Weiher, H. & Keller, W. *J. molec. Biol.* **180**, 577-600 (1984).
10. Schöler, H. R. & Gruss, P. *EMBO J.* **4**, 3005-3013 (1985).
11. Augereau, P. & Chambon, P. *EMBO J.* **5**, 1791-1797 (1986).
12. Schöler, H. R. & Gruss, P. *Cell* **36**, 403-411 (1984).
13. Mercola, M., Gorman, J., Mirell, C. & Calame, K. *Science* **227**, 266-270 (1985).
14. Zenke, M. *et al.* *EMBO J.* **5**, 387-397 (1986).
15. Weiher, H., Koenig, M. & Gruss, P. *Science* **219**, 626-631 (1983).
16. Ephrussi, A., Church, G. M., Tonegawa, S. & Gilbert, W. *Science* **227**, 134-140 (1985).
17. Wasylyk, C. & Wasylyk, B. *EMBO J.* **5**, 553-560 (1986).
18. Banerji, J., Olson, L. & Schaffner, W. *Cell* **33**, 729-740 (1983).
19. Wildeman, A. G. *et al.* in *Current Communications in Molecular Biology* (ed. Gluzman, Y.) 19-26 (Cold Spring Harbor Laboratory, New York, 1985).
20. Wildeman, A. G. *et al.* *Molec. cell. Biol.* **6**, 2098-2105 (1986).
21. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. *Cell* **33**, 717-728 (1983).
22. Mosthaf, L., Pawlita, M. & Gruss, P. *Nature* **315**, 597-600 (1985).
23. Sassone-Corsi, P., Wildeman, A. & Chambon, P. *Nature* **313**, 458-463 (1985).
24. Klein, G., Giovannella, B., Westman, A., Stehlin, J. S. & Munford, D. *Intervirology* **5**, 319-334 (1975).
25. Gidoni, D. *et al.* *Science* **230**, 511-517 (1985).
26. Kadonaga, J. T., Jones, K. A. & Tjian, R. *Trends biochem. Sci.* **11**, 20-23 (1986).
27. Gilbert, W., Maxam, A. M. & Mirzabekov, A. D. in *Control of Ribosome Synthesis* (eds Kjeldgaard, N. C. & Maaloe, O.) 139-143 (Munksgaard, Copenhagen, 1986).
28. Singh, H., Sen, R., Baltimore, D. & Sharp, P. A. *Nature* **319**, 154-158 (1986).
29. Toozé, J. (ed.) *DNA Tumor Viruses* (Cold Spring Harbor Laboratory, New York, 1982).
30. Barrera-Saldana, H. *et al.* *EMBO J.* **4**, 3839-3849 (1985).
31. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).

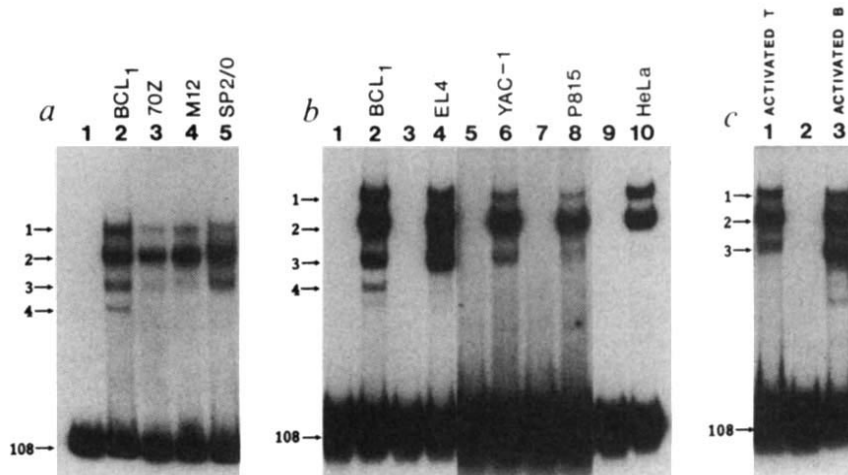
Interaction of cell-type-specific nuclear proteins with immunoglobulin V_H promoter region sequences

Nicholas F. Landolfi, J. Donald Capra & Philip W. Tucker

Department of Microbiology, Southwestern Medical School, University of Texas Health Science Center at Dallas, Dallas, Texas 75235, USA

All human and murine immunoglobulin heavy chain variable region (V_H) genes contain the sequence ATGCAAAT approximately 70 nucleotides 5' from the site of transcription initiation^{1,2}. This octanucleotide, in reverse orientation, is also found in all light chain variable region (V_L) genes^{1,2}, and in the immunoglobulin heavy chain transcriptional enhancer². Transfection studies have established that this octamer is involved in the lymphoid-specific transcription of immunoglobulin genes²⁻⁶. Octamer-containing fragments have been reported to bind a factor present in nuclear extracts of human cell lines; however, identical binding activity was detected in both B lymphoid and non-lymphoid cells⁷. Here we establish that nuclear extracts from distinct cell types differ

Fig. 3 Heterogeneous binding activity of nuclear extracts of various cell types to the octamer-containing fragment. The assays were performed as in Fig. 2. Migration of the arbitrarily-numbered protein-DNA complexes and the free fragment is indicated on the left of each panel. All extracts were prepared according to Dignam *et al.*²¹. None of these extracts exhibited binding activity to fragment 310 (data not shown). Different extract preparations of the same cell type exhibited consistent retention patterns with octamer-containing fragments. *a*, Autoradiogram of gel migration retardation assay examining the binding activity of nuclear extracts prepared from cell lines representing various stages of B-lymphocyte development. The source of each extract is indicated above the lane; lane 1 received no extract. The cell lines and their reported state of differentiation: 70Z, murine pre-B cell; M12, murine immature B cell; BCL₁, murine mature B cell; SP2/0, murine myeloma. *b*, Autoradiogram of gel migration retardation assay examining the binding activity of nuclear extract prepared from various differentiated cell types. The source of each extract is indicated above each lane; lanes 1, 3, 5, 7 and 9 received no extract. The cell lines and their characterization: BCL₁, mature murine B lymphocyte; EL4, murine thymoma; YAC-1, mature murine T lymphoma; P815, murine mastocytoma; HeLa, human carcinoma. *c*, Autoradiogram of gel migration retardation assay examining binding activity of nuclear extracts prepared from normal activated T- and B-lymphocyte populations. Activated B cells were prepared by incubating murine spleen cells for 3 days in the presence of lipopolysaccharide ($70 \mu\text{g ml}^{-1}$) and dextran sulphate ($20 \mu\text{g ml}^{-1}$)²²; activated T cells were prepared by incubating murine spleen cells in the presence of concanavalin A and maintaining the cells *in vitro* by addition of interleukin-2 as described²². Lane 2 received no extract.



tion patterns observed with these extracts were very similar to that of BCL₁ (for ease of discussion, the four protein-DNA complexes observed with the BCL₁ extract have been arbitrarily numbered 1-4), although quantitative differences could be noted (Fig. 3*a*). For example, species 3 was detected in lower abundance in the 70Z and M12 extracts than it was in the BCL₁ or SP2/0 extracts.

We next examined cell lines which were not of B-lymphocyte origin. Assays using fragment 108 and extracts prepared from EL4 (thymoma), YAC-1 (T-cell lymphoma), P815 (mastocytoma), and HeLa (carcinoma) cells established that there are qualitative as well as quantitative differences in the binding activity detected in these nuclear extracts (Fig. 3*b*). The migration-retarded species observed with the EL4 and YAC-1 extracts are similar, but differ quantitatively; species 1 and 3 were less abundant in the YAC-1 extract. Also, the EL4 extract exhibits an additional species that migrates between species 2 and 3. The two non-lymphoid extracts, P815 and HeLa, are also very similar; both exhibit species 1 and 2, however the P815 extract exhibits considerably less species 1 than does HeLa. Little or no species 3 is detected in either of these two extracts. Extracts were also prepared from mitogen-activated B lymphocytes and mitogen-activated T lymphocytes. On analysis with fragment 108 (Fig. 3*c*), the activated B-cell extract exhibited essentially equimolar amounts of species 2 and 3, while the activated T-cell extract exhibited species 1 and 2, but little or no species 3. These results demonstrate that nuclear extracts prepared from cells of distinct lineages display a heterogeneity in both the type and level of specific reactivity with the octamer-containing fragment.

The ability of the duplex 20-mer to inhibit competitively the formation of all four of the species in the BCL₁ pattern (Fig. 2*c*), as well as the two species detected in the HeLa cell extract (data not shown), implied that all sequences involved in the formation of these protein-DNA complexes resided within these 20 nucleotides. However, when the duplex 20-mer was end-labelled and used as a substrate in the assay, species 1 was not detected with either the BCL₁ or the HeLa extracts (Fig. 4, lanes 5 and 6). This protein-DNA complex was clearly evident when fragment 30 was used as a substrate (Fig. 4, lanes 2 and 3). Thus, although the duplex 20-mer could inhibit the formation of species 1, it was not a sufficient substrate for the detection of this species. Therefore, the formation of species 1 is dependent

upon both a sequence within the 10 nucleotides which comprise the 5' end of fragment 30 and a sequence located within the 20-mer. Species 1 may result from a single protein which has binding specificity for two separated DNA sequences, or it may result from a protein interacting with both DNA (the 5' 10 nucleotides of fragment 30) and another protein bound to a sequence within the 20-mer. Nevertheless, a sequence 5' to the octamer is involved in a protein-DNA interaction.

The results reported here differ from those reported by Singh *et al.*⁷, who have described the interaction of a human factor, IgNF-A, with the octamer-containing fragments of the promoter region of murine V_H and V_L genes, as well as with the immunoglobulin heavy chain enhancer. Singh *et al.* observed only two migration-retarded species, and reported no cell-type specific difference in the pattern of protein-DNA complexes, as extracts prepared from a human B lymphoma (EW) and a human carcinoma line (HeLa) exhibited identical reactivity⁷. We also observed two protein-DNA complexes on analysing a HeLa cell extract in the gel retention assay. However, using lymphoid cell extracts, we observed up to four distinct protein-DNA complexes. The reason for this discrepancy is unclear, and may simply reflect technical differences in extract preparation and/or assay conditions. Extract mixing experiments have established that the additional, faster-migrating species (such as species 3 and 4) which we observe with lymphoid extracts are not an artefact of extract contamination with cellular proteases (data not shown). Because all the extracts examined in our study, with the exception of HeLa, were of murine origin, it is possible that the additional complexes we observed are unique to this species.

Our analysis has not identified a protein unique to B lymphocytes which possesses octamer-binding activity. However, the protein responsible for species 3 displays a pattern of expression consistent with a role in the lymphoid-specificity of immunoglobulin expression. Species 3 is not detected in non-lymphoid cell lines, but is detected in varying levels in all the lymphoid extracts, and thus is lymphoid specific. The relative level of species 3 is higher in the more mature B-cell lines (BCL₁ and SP2/0) than in the immature lines (70Z and M12). Species 3 is most predominant in the extract of mitogen-activated B lymphocytes, a population which exhibits an up to 10-fold higher level of immunoglobulin transcription than neoplastic B

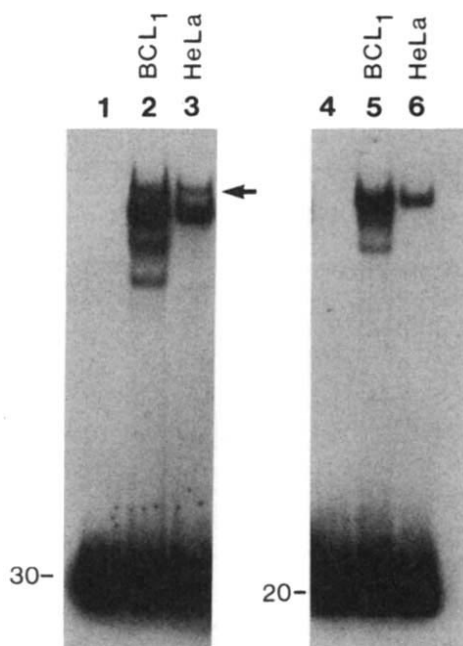


Fig. 4 The slowest-migrating species (species 1) is dependent upon nucleotides 5' of the octamer. Autoradiogram of gel migration retardation assay examining binding activity of BCL₁ and HeLa extracts on fragment 30 (lanes 2, 3) and the duplex 20-mer (lanes 5 and 6) which is the twenty 3' nucleotides of fragment 30. The assay was carried out as in Fig. 2. Species 1 is detected with fragment 30 (arrow, lanes 2 and 3), but is not detected with the duplex 20-mer (lanes 5 and 6).

cells^{11,12}. The presence of species 3 in the EL4 extract at high levels would seem contradictory; however this line produces abundant 'sterile' μ transcripts¹³. These 'sterile' transcripts initiate 43–86 bases downstream of the enhancer-associated octamer sequence within the J_H–C _{μ} intron^{14,15}. Thus, it is possible that the enhancer-associated octamer and the protein responsible for species 3 are involved in the initiation of the μ transcripts detected in EL4. Therefore, species 3 may represent the protein–DNA interaction involved in the octamer-associated lymphoid-specific transcription of immunoglobulin genes.

The basis of the pattern of fragment migration retardation (up to four distinct species) is unclear. Each may represent the interaction of a distinct nuclear protein with a specific DNA sequence. Alternatively, each species may represent multimers of a single factor bound to the DNA. A single protein, differentially post-translationally modified (for example, phosphorylated) could be responsible for generating some of the species. It is also possible that one or more of these complexes represent the interaction of one protein with another protein which is bound to the DNA; the putative second protein need not possess sequence-specific DNA binding capability, and may have affinity for the other nuclear factor only when the latter is bound to DNA. There is no reason to believe that only one of the above mechanisms is responsible for the formation of all the DNA–protein complexes; it is more likely that the pattern observed is the result of a combination of these mechanisms.

The presence of species which exhibit an identical amount of migration retardation with different nuclear extracts (for example species 2 which is detected in all extracts examined) does not necessarily indicate the presence of identical factors. Equivalent migration inhibition may only indicate a basic biochemical similarity among certain nuclear proteins (relative molecular mass and/or charge). It is possible that different cell types contain biochemically similar nuclear factors which have the identical DNA sequence specificity, but differ in their effector function. Such factors may behave identically in the gel retention assay.

Our results indicate a wide tissue distribution of nuclear proteins capable of interacting with octamer-containing fragments. In fact, the octamer motif is not limited to immunoglobulin-associated promoter and enhancer regions. The sequence ATGPyAAAT is located upstream of human and *Xenopus* U1 and U2 small nuclear RNA genes and is required for accurate transcription^{16–19}. This suggests that the octamer sequence element is involved in transcriptional control of a number of eukaryotic genes.

We thank Sherri McGehee, Maureen Willard, Chhaya Das, Caroline Humphries and Madelaine Allan for technical assistance, and Marilyn Gardner for preparation of this manuscript. N.F.L. is a fellow of the Leukemia Society of America. This work was supported by NIH research grants AI-18016, GM-1751 and GM-31689.

Received 28 May; accepted 30 July 1986.

1. Parslow, T. G., Blair, D. L., Murphy, N. J. & Granner, D. K. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2650–2654 (1984).
2. Falkner, F. G. & Zachau, H. G. *Nature* **310**, 71–74 (1984).
3. Bergman, Y., Rice, D., Grosschedl, R. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7041–7045 (1984).
4. Mason, J. O., Williams, G. T. & Newberger, M. S. *Cell* **41**, 479–487 (1985).
5. Grosschedl, R. & Baltimore, D. *Cell* **41**, 885–897 (1985).
6. Foster, J., Stafford, J. & Queen, C. *Nature* **315**, 423–425 (1985).
7. Singh, H., Sen, R., Baltimore, D. & Sharp, P. *Nature* **319**, 154–158 (1986).
8. Gronowicz, E. S., Doss, C. A., Howard, F. D., Morrison, D. C. & Strober, S. *J. Immun.* **125**, 976–980 (1980).
9. Garner, M. M. & Revzin, A. *Nucleic Acids Res.* **9**, 3047–3060 (1981).
10. Fried, M. & Crothers, D. M. *Nucleic Acids Res.* **9**, 6505–6525 (1981).
11. Yuan, D., Gilliam, A. C. & Tucker, P. W. *Fedn Proc.* **44**, 2652–2659 (1985).
12. Yuan, D. & Tucker, P. W. *J. exp. Med.* **160**, 564–583 (1984).
13. Kemp, D. J., Morahan, G., Cowman, A. F. & Harris, A. W. *Nature* **301**, 84–86 (1983).
14. Nelson, K. J., Haimovich, J. & Perry, R. P. *Molec. cell Biol.* **3**, 1317–1332 (1983).
15. Lennon, G. G. & Perry, R. P. *Nature* **318**, 475–478 (1985).
16. Ares, M. Jr, Mangin, M. & Weiner, A. M. *Molec. cell Biol.* **5**, 1560–1570 (1985).
17. Mattaj, J. W., Leinhardt, S., Jiricny, J. & DeRobertis, E. M. *Nature* **316**, 163–167 (1985).
18. Krol, A., Lund, E. & Dahlberg, J. E. *EMBO J.* **4**, 1529–1535 (1985).
19. Ciliberto, G., Buckland, R., Cortese, R. & Philipson, L. *EMBO J.* **4**, 1537–1543 (1985).
20. Knapp, M. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 2996–3000 (1982).
21. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
22. Landolfi, N. F. & Cook, R. G. *Molec. Immun.* **23**, 297–309 (1986).

Light-inducible and tissue-specific pea *lhcp* gene expression involves an upstream element combining enhancer- and silencer-like properties

J. Simpson*, J. Schell, M. Van Montagu & L. Herrera-Estrella*

Laboratorium voor Genetica, Rijksuniversiteit Gent, B-9000 Gent, Belgium

Studies in viral and mammalian cell systems have identified regulatory elements, collectively termed enhancers, which are *cis*-acting elements effective in either orientation and at both 5' and 3' ends of a gene, and which increase transcription when linked to either their own or a heterologous promoter gene system^{1–5}. Certain upstream elements in yeast have also been shown to exert a negative effect on the rate of transcription of cell-cycle-specific genes and these have been called silencers⁶. Two of the best characterized plant gene families are those of the small subunit of the ribulose 1,5-bisphosphate carboxylase (*rbcS*) genes⁷ and the light-harvesting chlorophyll *a/b*-binding protein (*lhcp*) genes⁸. The first enhancer element in plants has recently been described showing that an upstream element of an *rbcS* gene could, in both orientations, confer light-inducible expression on a heterologous promoter/gene system⁹. We report here that a 247-base pair (bp) element from an *lhcp* gene acts not only as a light-inducible enhancer but also as a tissue-specific 'silencer'.

* Present address: Centro de Investigación y Estudios Avanzados, Apdo Postal 629, 26500 Irapuato, Gto, Mexico.

The *rbcs*^{10,11} and *lhcp*^{8,12} genes are found in small multi-gene families in the nuclear genome of plant cells, although their gene products are active in the chloroplasts¹³⁻¹⁵. They are light-inducible (see ref. 16 for a review) and under phytochrome control^{16,17}, and both *rbcs*^{18,19} and *lhcp*^{20,21} genes are expressed only in specific plant tissues. When a 400-bp fragment of the *lhcp* gene *AB80* (ref. 22) was fused to the NPT(II) (neomycin phosphotransferase II) coding sequence, the chimaeric gene construct (pPC in Fig. 1) was shown to be regulated in a light-inducible and tissue-specific fashion in transgenic plants²⁰. Deletion of 247 bp (-100 to -347) of this 400-bp sequence, leaving the TATA and CAAT boxes intact, resulted in loss of resistance to kanamycin (J.S., unpublished observations). To determine whether the 247-bp sequence contained regulatory elements involved in light-inducible or tissue-specific expression, this sequence was cloned in both orientations upstream of the constitutively expressed nopaline synthase (*nos*) promoter and fused to the NPT(II) coding sequence (Fig. 1). The resulting constructs were pABN1, with the upstream element (UE) in the same orientation to the *nos* promoter as originally found in the 5'-flanking region of the *AB80* gene; pABN2, with the upstream element in the opposite orientation to pABN1; and pABN3, with the upstream element in tandem in the same orientation as pABN1 (see Fig. 2).

Regenerated plants transformed by pABN1, pABN2, pABN3 or the control pLGV1103neo were obtained by *Agrobacterium*-mediated transformation and plasmids containing the chimaeric constructs were conjugated²³ from *Escherichia coli* to an *Agrobacterium* strain harbouring a non-oncogenic Ti plasmid derivative, pGV3850 (ref. 24), in which DNA sequences of interest can be easily inserted between flanking border repeats of the T-DNA by a single recombination event. *Agrobacterium* transconjugants carrying co-integrates between pGV3850 and pABN1, pABN2 or pABN3 were selected by plating in medium containing kanamycin and used to infect SRI tobacco leaf disks²⁵. Transformed tissue was selected on 75 $\mu\text{g ml}^{-1}$ kanamycin and transformed shoots were verified by assaying for nopaline synthase activity²⁶. These shoots were then transferred to vermiculite and grown to maturity. pABN1-, pABN2- and pABN3-transformed shoots appeared faster on kanamycin-containing medium than did shoots transformed with the control pLGV1103neo construct (Figs 1, 2) involving only the *nos* promoter fused to the NPT(II) gene. Moreover, these shoots failed to form roots on kanamycin-containing medium whereas pLGV1103neo-transformed shoots formed roots normally.

To determine whether the presence of the *AB80* upstream element²² could influence expression of the *nos*-NPT(II) construct, transgenic plants were examined for light inducibility of NPT(II) activity. Plants were grown for 5 days in complete darkness, the level of NPT(II) activity was determined and they were then placed in continuous light conditions for 4 days after which NPT(II) in leaves of the same size and position on the plant was again determined. Alternatively, transformed plants of each type were divided in two, eventually giving rise to two plants with identical genomes. One set of these plants was kept in light while the other set was placed in darkness for 5 days. Leaf extracts containing equal amounts of proteins were then assayed for NPT(II) activity or extracted for hybridization analysis.

These experiments showed that the level of NPT(II) activity in extracts of pLGV1103neo transgenic plants was unaffected by the light conditions, whereas pABN1, pABN2 and pABN3 transgenic plants showed four to eight times higher NPT(II) activity in the light than under dark growth conditions. pABN3 plants were estimated to have 10-13 times higher activity when grown in light than in the dark (Fig. 2). Three independent transgenic plants have been tested for each construct and all show similar results.

To confirm that the observed light inducibility of pABN1, pABN2 and pABN3 transgenic plants reflected changes in the

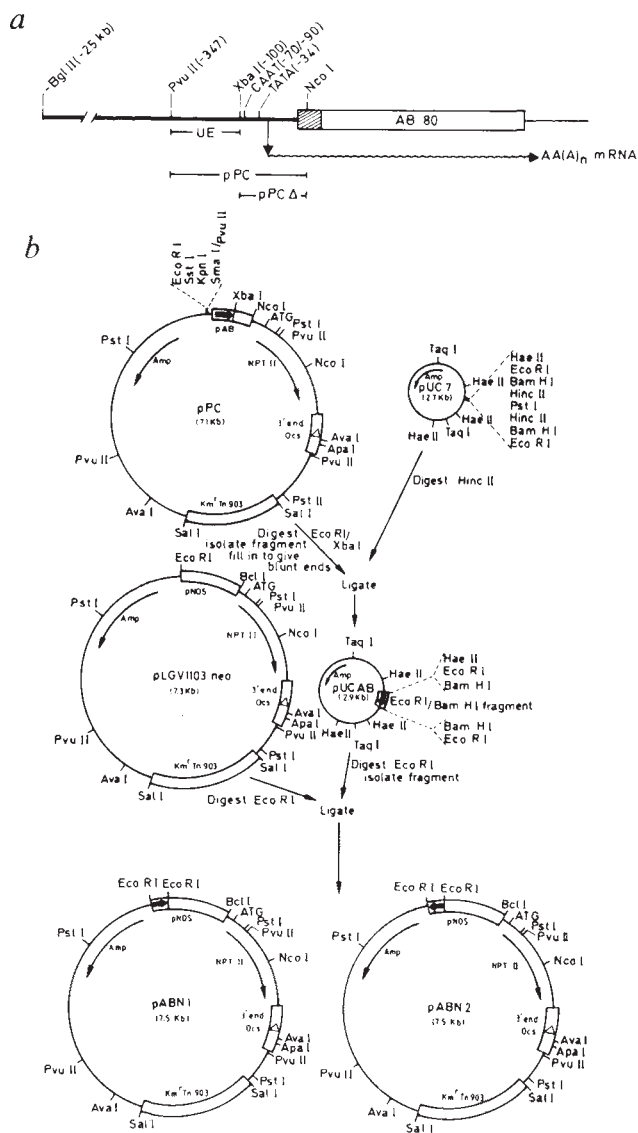


Fig. 1 Diagram of the formation of chimaeric constructs with the *AB80* upstream element inserted in both orientations 5' to the *nos*-NPT(II) construct. *a*, Representation of the *AB80* gene including the position of the TATA and CAAT boxes, as well as relevant restriction sites. The 5'-flanking region contained in the pPC chimaeric gene and its inactive deletion (pPCΔ) are shown below the *AB80* gene. The hatched box represents the region encoding the transit peptide of the LHCP polypeptide. *b*, Chimaeric gene constructs were made using standard DNA techniques as described previously³⁴. An EcoRI-XbaI fragment was isolated from pPC and the ends were filled in using Klenow polymerase and the blunt-ended fragment was then cloned into the HincII-digested pUC7 vector. An EcoRI fragment from the construct pUCAB was isolated and cloned into the EcoRI-digested pLGV1103neo. Clones containing the upstream element in both orientations, as found in the *AB80* gene (pABN1) and in the opposite orientation (pABN2), and a third clone containing a tandem insert of the *AB80* upstream in the same orientation as pABN1, were purified and conjugated to the *Agrobacterium* strain pGV3850 for subsequent transformation of plants. Kb, kilobases.

level of steady-state messenger RNA, total RNA from the pooled leaves of ten transgenic plants of each type was extracted from light- or dark-grown plants. The level of NPT(II) mRNA, determined by dot-blot hybridization, was approximately 5-fold higher in leaves of light-grown pABN1 and pABN2 transgenic plants than in the pLGV1103neo control, and about 10-fold higher in leaves of pABN3 transgenic plants than in those grown in the dark. These experiments show that the presence of the *AB80* upstream element confers light-inducible properties on

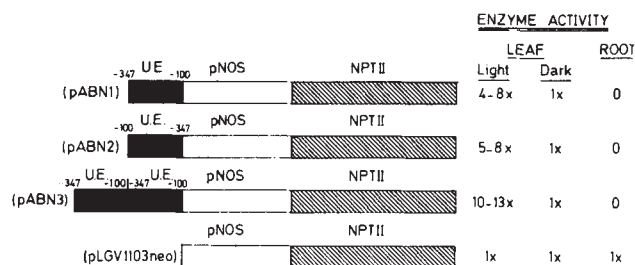


Fig. 2 Diagram of constructs showing relative NPT(II) activities under light and dark conditions, and in leaf and root tissue. The positioning and orientation of the *AB80* upstream element in pABN1, pABN2 and pABN3 are shown. The solid black box represents the upstream element, the open box the *nos* promoter and the hatched box the NPT(II) coding sequence. Relative NPT(II) activities in light- and dark-grown plants, and in leaf and root tissue, are presented. Similar results were obtained by quantitating mRNA dot-blots. The NPT(II) activity for the pLGV1103neo construct alone is taken as 1, and the pABN1, pABN2 and pABN3 activities are expressed relative to this. NPT(II) activity was determined in leaf and root extracts of transgenic plants by the *in situ* detection method on non-denaturing polyacrylamide gels^{35,36} using equal amounts of protein from all extracts. Quantitation was carried out by measuring the kanamycin-³²P₄ bound to the phosphocellulose paper in a liquid scintillation counter and relating this to the amount of total protein used in the assay. To quantitate mRNA, the ³²P-labelled dots were cut from the filter and put through a scintillation counter, and the results using RNA extracted from light- and dark-grown plants or leaf and root tissue compared.

the *nos-NPT(II)* gene and that two copies of the upstream element 5' to the *nos* promoter seem to have an additive effect.

To demonstrate that the increase in the level of NPT(II) mRNA on illumination of pABN1 and pABN2 transgenic plants corresponds to an increase in transcription initiated at the same point as in the control pLGV1103neo construct, RNA from each type of transgenic plant was subjected to S₁ mapping analysis (Fig. 3b). A single band of ~276 nucleotides was protected against S₁ nuclease using RNA from pLGV1103neo transgenic plants. This corresponds to the 5' end of the mRNA previously observed for the *nos* gene itself²⁷, and this same band was observed when RNA from light- and dark-grown pABN1 and pABN2 plants was analysed. The 247-bp *AB80* upstream element sequence therefore increases the level of transcripts initiated at the same site as in pLGV1103neo transgenic plants.

To determine whether the *AB80* upstream element could also regulate the *nos-NPT(II)* fusion in a tissue-specific manner, extracts from leaves and roots of pABN1, pABN2 and pABN3 transgenic plants, a pLGV1103 control plant and a plant containing the intact *AB80* promoter fused to the NPT(II) gene were examined for NPT(II) activity. The results (Fig. 4a) are quite striking, for whereas the control *nos-NPT(II)* construct is expressed at a similar level in both leaf and root tissue, the pABN1, pABN2 and pABN3 constructs show strong expression in leaf tissue but no, or very low, expression in roots, mirroring the expression pattern of the intact *AB80* promoter (Fig. 4a, lanes 1, 2). Three or four transgenic plants of each type (pABN1, pABN2, and pABN3) have been tested and all show similar results. Again, to confirm that the NPT(II) activity reflects events occurring at the transcription level, total mRNA was isolated from the pooled leaf or root tissue of two transgenic plants of each type and the level of NPT(II) mRNA was determined by dot-blot hybridization. Figure 4b shows that the NPT(II) mRNA levels correspond to results observed from the NPT(II) assays. Thus, the *AB80* upstream element not only has enhancing activity in light-grown plant tissue but also has silencing activity in the roots of transgenic plants.

Our results show that the enhancer activity resides within a 247-bp fragment from -100 to -347 bp in the 5'-flanking region

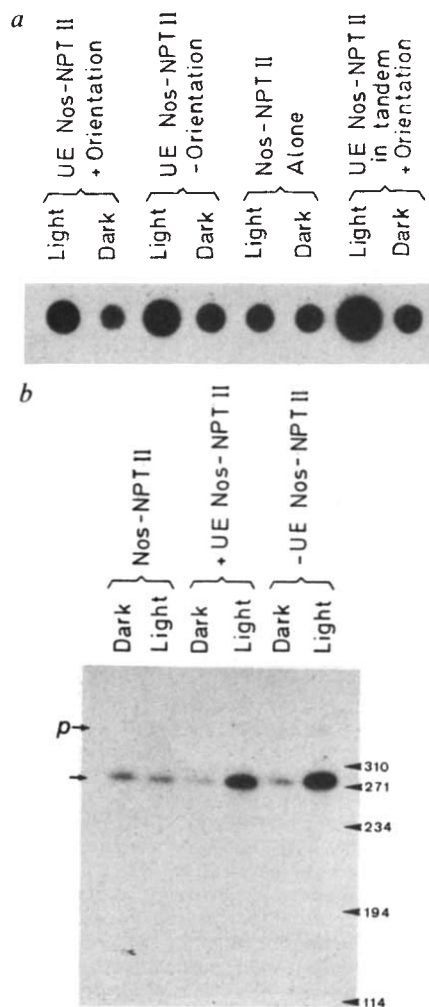


Fig. 3 a, Comparison of NPT(II) mRNA present in light- and dark-grown plants harbouring the pABN1 and pABN2 constructs. b, S₁ mapping of mRNA transcripts present in light- and dark-grown plants harbouring the pABN1 and pABN2 constructs. The figure compares hybridization/protection studies following S₁ nuclease digestion of mRNA from pABN1, pABN2 and pLGV1103neo transgenic plants. The size of the protected fragments was estimated by reference to the mobility of *Hae*III-digested ΦX174 DNA markers. The arrow indicates the mobility of the 275-bp protected fragment, and p the mobility of the complete single-stranded probe.

Methods. a, Leaves of SR1 tobacco plants were cut into 1-cm² pieces and floated upside-down in K3 medium containing 0.2 mg l⁻¹ benzyl aminopurine, 0.1 mg l⁻¹ naphthalene acetic acid. Aliquots (50 μl) of an undiluted *Agrobacterium* culture grown overnight in minimal A medium were then added. Three days later, the leaf pieces were washed in K3 containing 500 μg ml⁻¹ Claforan (Hoechst) and transferred to solid Linsmaier and Skoog medium containing 50 μg ml⁻¹ kanamycin sulphate (Sigma), 500 μg ml⁻¹ Claforan and hormones (1 mg l⁻¹ BAP) to stimulate shoot formation. When shoots reached ~0.5 cm they were transferred to media without hormones but containing kanamycin and Claforan as before. Shoots remaining green after 2-3 weeks were tested for nopaline synthase activity²⁶. Positive shoots were then transferred to non-selective media to allow root formation. Total RNA was prepared separately from the leaves of light- and dark-grown pABN1-, pABN2- and pLGV1103neo-transformed plants by a technique described elsewhere³². Total RNA (20 μg) from each organ of both plants was spotted onto nylon filters using the minifold device³⁷. A DNA fragment containing the coding sequence of the NPT(II) gene was labelled to 1 × 10⁸ d.p.m. μg⁻¹ by nick translation with ³²P-dCTP and used as a hybridization probe against the RNA spots. b, S₁ mapping was carried out³⁷ using 20 μg of poly(A)⁺ RNA and a single 430-bp strand probe spanning the *nos* promoter and 275 bp into the NPT(II) coding sequence from the start point of transcription of *nos* promoter.

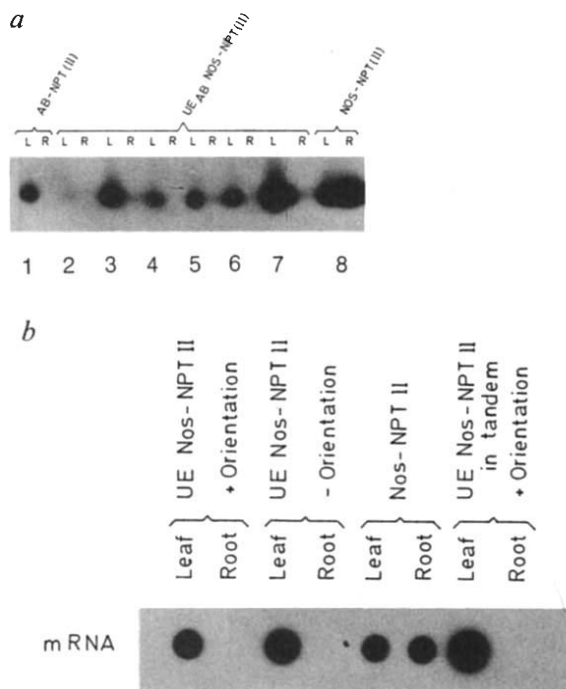


Fig. 4 Comparison of NPT(II) activity and mRNA present in leaves and roots of pABN1-, pABN2-, pABN3- and pLGV1103-transformed plants. *a*, NPT(II) activity was determined as for Fig. 2. The figure shows the NPT(II) activity present in transgenic plants containing the NPT(II) coding sequence under control of the *AB80* promoter (*AB-NPT(II)*), the *nos* promoter (*nos-NPT(II)*). The different constructions containing the *AB80* upstream element 5' to the *nos* promoter are shown as follows: lanes 2 and 3, upstream element in the same orientation as found in the *AB80* gene; lanes 4 and 5, upstream element in the opposite orientation; lanes 6 and 7, two copies of the upstream element in tandem in the same orientation as in the *AB80* gene. *b*, The level of NPT(II) transcription was estimated in leaves and roots of transgenic plants by dot-blot hybridization as described for Fig. 3. The hybridization corresponding to each of the tissues from the transgenic plant containing the different constructions is indicated.

of the *AB80* gene. The S_1 mapping experiment demonstrates that the enhanced transcription in light conditions is initiated at the same point as in the *nos-NPT(II)* construct lacking the *AB80* upstream element, suggesting that this element is acting on the *nos* promoter and not as an independent promoter sequence. The initial selection for transformed tissue seems to reflect these results. Deletion of the upstream element from the 400-bp intact *AB80* promoter fragment resulted in loss of kanamycin resistance, probably due to very low or no activity of the remaining promoter sequences, while the finding that pABN1, pABN2 and pABN3 shoots appear more quickly than pLGV1103neo shoots is probably due to the increased NPT(II) activity observed in these shoots compared with that in the pLGV1103neo shoots. Experiments placing the upstream element at the 3' end of the *nos-NPT(II)* chimaeric gene have produced no kanamycin-resistant plant tissue following several transformation attempts, perhaps because only the suppressive effect is functioning in this location. Moreover, lack of root formation on kanamycin-containing medium for pABN1, pABN2 and pABN3 constructs correlates with the lack of NPT(II) activity found in the roots of mature plants.

The most interesting result of this work is that the *AB80* upstream element both enhances expression in the light and silences expression in the roots of transgenic plants of a gene which is normally expressed in this tissue. Both the enhancer and silencing effects are independent of the orientation of the upstream element. Any model for the regulation of the *AB80*

upstream element must take this orientation independence into account, together with the fact that the distance to the start point of transcription is not restrictive, for in the *nos-NPT(II)* construct the element is located from -463 to -733 bp, whereas the normal position in the *AB80* 5'-flanking sequence is between -100 and -347 bp²². Furthermore, other factors are known to influence expression of endogenous *lhcp* genes, including the involvement of chloroplast development²⁸⁻³³. One model to explain the silencing effect of this element in roots is that in the absence of developed chloroplasts some factor(s) acting on the upstream element prevent transcription; this would be the case in root tissue. In photosynthetic tissue, however, we propose that during development from proplastid to chloroplast, this factor becomes inactivated, allowing a second positively acting factor that probably interacts with phytochrome to act on the upstream element, thereby enhancing transcription upon illumination.

This undoubtedly oversimplifies the regulatory mechanism and other elements within the 5'-flanking sequence of the *AB* are probably involved in controlling light-inducible and tissue-specific expression. This is supported by the fact that fragments greater than 400 bp of the *AB80* 5'-flanking sequence increase expression of chimaeric *AB-NPT(II)* fusions²¹. The hypothesis of multiple elements with the same function is supported by the finding that the pABN3 construct with the upstream element in tandem gives higher levels of expression than pABN1 and pABN2, suggesting that plant gene regulation may involve the choice between the action of positive or negative regulators mediated by enhancer and silencer elements, in a fashion similar to those found in prokaryotic organisms mediated by much simpler mechanisms of regulation.

We thank Allan Caplan, Patrick Dhaese and Gilbert Engler for critical reading of the manuscript, Martine De Cock for typing, and Karel Spruyt and Albert Verstraete for figures. This work was supported by grants of the 'ASLK-Kankerfonds', the 'Fonds voor Wetenschappelijk Geneeskundig Onderzoek' (FGWO 3.0001.82) and the Services of the Prime Minister (OOA 12.0561.84). J.S. thanks CEC for a predoctoral fellowship.

Received 30 April; accepted 31 July 1986.

1. Banerji, J., Rusconi, S. & Schaffner, W. *Cell* **27**, 299-308 (1981).
2. Moreau, P. *et al. Nucleic Acids Res.* **9**, 6047-6068 (1981).
3. Fromm, M. & Berg, P. *Molec. cell. Biol.* **3**, 991-999 (1983).
4. Banerji, J., Olson, L. & Schaffner, W. *Cell* **33**, 729-740 (1983).
5. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. *Cell* **33**, 717-728 (1983).
6. Brand, A. H., Breiden, L., Abraham, J., Sternglanz, R. & Nasmyth, K. *Cell* **41**, 41-48 (1985).
7. Cashmore, A. R. in *Genetic Engineering of Plants—An Agricultural Perspective* (eds Kosuge, T., Meredith, C. P. & Hollaender, A.) 29-38 (Plenum, New York, 1983).
8. Timko, M. P. & Cashmore, A. R. *UCLA Symp. molec. cell. Biol. New Ser.* **12**, 403-412 (1983).
9. Timko, M. P. *et al. Nature* **318**, 579-582 (1985).
10. Dean, C., van den Elzen, P., Tamaki, S., Dunsmuir, P. & Bedbrook, J. *EMBO J.* **4**, 3055-3061 (1985).
11. Cashmore, A. R. *Cell* **17**, 383-388 (1979).
12. Coruzzi, G., Broglie, R., Cashmore, A. R. & Chua, N.-H. *J. biol. Chem.* **258**, 1399-1402 (1983).
13. Schmidt, G. W., Bartlett, S. G., Grossman, A. R., Cashmore, A. R. & Chua, N.-H. *J. Cell Biol.* **91**, 468-478 (1981).
14. Highfield, P. E. & Ellis, R. J. *Nature* **271**, 420-424 (1978).
15. Chua, N.-H. & Schmidt, G. W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6110-6114 (1978).
16. Thompson, W. F., Kaufman, L. S. & Watson, J. C. *BioEssays* **3**, 153-159 (1985).
17. Tobin, E. M. *Pl. molec. Biol.* **1**, 35-51 (1981).
18. Coruzzi, G., Broglie, R., Edwards, C. & Chua, N.-H. *EMBO J.* **3**, 1671-1679 (1984).
19. Herrera-Estrella, L. *et al. Nature* **310**, 115-120 (1984).
20. Lampka, G., Nagy, F. & Chua, N.-H. *Nature* **316**, 750-752 (1985).
21. Simpson, J. *et al. EMBO J.* **4**, 2723-2729 (1985).
22. Cashmore, A. R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2960-2964 (1984).
23. Van Haute, E. *et al. EMBO J.* **2**, 411-418 (1983).
24. Zambryski, P. *et al. EMBO J.* **2**, 2143-2150 (1983).
25. Horsch, R. B. *et al. Science* **227**, 1229-1231 (1985).
26. Otten, L. A. & Schilperoort, R. A. *Biochim. biophys. Acta* **527**, 497-500 (1978).
27. Gorman, C. M., Rigby, P. W. J. & Lane, D. P. *Cell* **42**, 519-526 (1985).
28. Müller, M., Viro, M., Balke, C. & Klopffsch, K. *Planta* **148**, 444-447 (1980).
29. Reiss, T., Bergfeld, R., Link, G., Thien, W. & Mohr, H. *Planta* **159**, 518-528 (1983).
30. Mayfield, S. P. & Taylor, W. C. *Eur. J. Biochem.* **144**, 79-84 (1984).
31. Oelmüller, R. & Mohr, H. *Planta* **167**, 106-113 (1986).
32. Eckes, P., Schell, J. & Willmitzer, L. *Molec. gen. Genet.* **199**, 216-224 (1985).
33. Simpson, J., Van Montagu, M. & Herrera-Estrella, L. *Science* **233**, 34-38 (1986).
34. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
35. Van den Broeck, G. *et al. Nature* **313**, 358-363 (1985).
36. Reiss, B., Sprengel, R., Will, H. & Schaller, H. *Gene* **30**, 211-218 (1984).
37. Bresser, J., Doering, J. & Gillespie, D. *DNA* **2**, 243-254 (1983).

Phorbol ester induces the transcriptional stimulatory activity of the SV40 enhancer

Richard J. Imbra* & Michael Karin

Division of Pharmacology, School of Medicine, M-013 H, University of California, San Diego, La Jolla, California 92093, USA

Phorbol ester tumour promoters can induce the transcription of a number of genes¹, including *c-myc* and *c-fos*². These genes are part of a group referred to as 'competence' genes^{3,4} because they are expressed very early after quiescent cells are stimulated to enter the cell cycle⁵⁻⁷. The 'competence' genes are coordinately induced by serum and by factors such as platelet-derived growth factor and fibroblast growth factor^{1,2,8-12}. These factors, as well as the tumour promoter, 12-*O*-tetradecanoyl phorbol-13-acetate (TPA), are thought to exert their action by a mechanism involving the activation of protein kinase C¹³⁻¹⁶. It is likely that these factors induce the transcription of the 'competence' genes either by activating specific transcription factors or by increasing their intracellular concentration; either mechanism may be mediated by protein kinase C. One approach to identifying such a putative transcription factor is to characterize the *cis*-acting transcriptional control elements that serve as a target site for the factor. Here we report that, in a human hepatoma cell line, TPA can specifically induce the activity of the simian virus 40 (SV40) transcriptional enhancer element. Since the SV40 enhancer is a thoroughly characterized *cis*-acting element¹⁷⁻¹⁹, this system may facilitate the eventual identification of the *trans*-acting factor(s) whose activity is modified by TPA treatment.

Recently we have found that expression of the human metallothionein II_A (*hMT-II_A*) gene is regulated at the transcriptional level by serum factors and by TPA. The human hepatoma cell line, HepG2, is especially useful for investigating this regulation, but expression of *hMT-II_A* is also regulated in a similar manner by these agents in primary human fibroblasts (R.J.I. and M.K., in preparation). To gain a better understanding of the mechanism by which TPA regulates expression of *hMT-II_A*, we performed transient expression experiments using the plasmid *phMT-II_A-CAT*, which contains the *hMT-II_A* promoter sequence linked to the transcription unit coding for the enzyme chloramphenicol acetyltransferase (CAT)²⁰. As a control for the *phMT-II_A-CAT* fusion, we used the plasmid *pSV2-CAT*, which contains the SV40 early promoter and enhancer elements linked to the CAT sequence²¹. Both these plasmids were introduced into HepG2 cells by the calcium phosphate technique²². After transfection was complete (4 h), the cells were incubated for 12–14 h in either normal growth medium, which contains 10% fetal calf serum (FCS), or in medium containing 0.5% FCS. To test the effect of inducers on *phMT-II_A-CAT* and *pSV2-CAT* expression, cells were incubated in the presence of either 5×10^{-6} M CdCl₂ or 100 ng ml⁻¹ TPA. Surprisingly, expression of CAT activity from *pSV2-CAT* was increased up to 45-fold by treatment with TPA, while expression of *phMT-II_A-CAT* is increased 2- to 3-fold only in cells incubated in low serum (Fig. 1). In contrast, expression of *pSV2-CAT* was not affected by the presence of CdCl₂, whereas expression of *phMT-II_A-CAT* increased 20-fold. The results in Fig. 1 also demonstrate that the basal level of *pSV2-CAT* expression in HepG2 cells is serum-dependent and decreases 5- and 6-fold following incubation in low serum medium. This effect appears to be specific to *pSV2-CAT*, because expression of *phMT-II_A-CAT* is only slightly decreased following incubation in low serum medium. The maximal level of expression of *pSV2-CAT* after TPA treatment is similar in cells incubated in low or high serum, suggesting that the level of

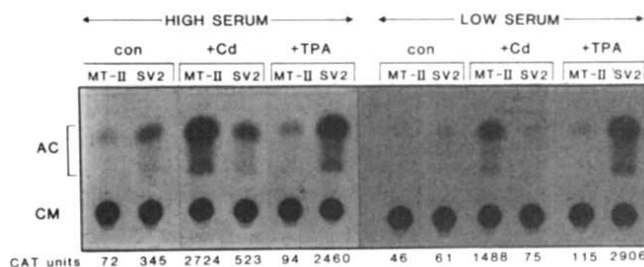


Fig. 1 Expression of the SV40-CAT transcription unit is induced by TPA. HepG2 cells were transfected with *pSV2-CAT* and *phMT-II_A* and incubated in either control medium (con), or in the presence of 10 μ M CdCl₂ or 100 ng ml⁻¹ of TPA for 12 h, at the end of which the cells were collected and the level of CAT expression determined. 'Low serum' indicates that cells were kept in Dulbecco's modified Eagle's medium (DMEM) containing 0.5% fetal calf serum (FCS) after the transfection, and 'High serum' indicates that cells were kept in their normal growth medium (DMEM + 10% FCS). Shown are the results of a typical CAT assay.

Methods. HepG2 cells were seeded at approximately 10^6 cells per 100 mm plate and transfected 24 h later with 2 μ g of plasmid DNA, using the calcium phosphate precipitation method²². Incubation with precipitate was for 4 h followed by a 2 min incubation in HEPES-buffered saline containing 15% glycerol. One-half of the cells received fresh medium containing 10% FCS (High serum) and the other half medium with 0.5% FCS (Low serum). At this point the different inducers were added at the indicated concentrations. Cells were collected 12–14 h later and CAT activity in cell extracts was determined as described^{21,22}. Results were quantified by cutting out the bands containing unreacted ¹⁴C-chloramphenicol (CM) and its acetylated forms (AC) and determining the amount of radioactivity by liquid scintillation counting. The levels of CAT expression in U per μ g protein are indicated above each lane.

serum does not affect gene expression non-specifically by influencing the transfection efficiency.

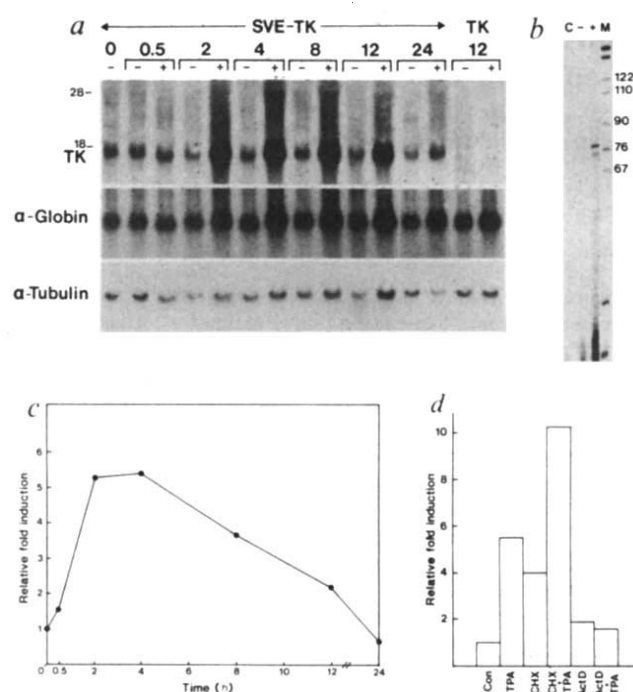
To characterize the unexpected effect of TPA on *pSV2-CAT* expression, we first determined whether the induction effect was dependent on the amount of transfected plasmid. The results in Table 1 show that the extent of induction is dependent on the amount of transfected DNA; induction varies from 20-fold at low DNA input to 3- to 5-fold at high DNA input. Furthermore, in the presence of TPA, the maximal induced level of CAT expression is reached at intermediate input DNA levels while the basal level of expression continues to increase at the highest DNA concentration used. This dependency on DNA input of the TPA response is different from that observed earlier for Cd²⁺ induction of *phMT-II_A-CAT*, which was independent of the amount of transfected DNA²⁰. This could be explained by a *trans*-acting factor that mediates the TPA effect and which is present in a limiting amount.

To examine further the promoter specificity of the TPA effect, transfection experiments were performed using several other cellular and viral promoters. Only vectors containing the SV40 enhancer show significantly elevated expression after TPA treatment (Table 2). None of the other viral or cellular promoters tested exhibited any significant response. We also observed that induction of *pSV2-CAT* by TPA occurs in HepG2 cells, Fao cells (a rat hepatoma cell line), and 293 cells (adenovirus-transformed human kidney cells), but not in HeLa, GC (a pituitary cell line) or Hep3B cells (a different human hepatoma cell line) (data not shown). Since the response was greatest using HepG2 cells, these cells were used in subsequent experiments. Furthermore, induction of *pSV2-CAT* expression was observed with TPA, but not with 4- α -phorbol-12,13-didecanoate (data not shown), a phorbol ester that is not a tumour promoter or an activator of protein kinase C (ref. 23).

The induction response of several transcription units has been reported to be due to the presence of inducible enhancer elements. For example, murine mammary tumour virus^{24,25} and *hMT-II_A* (ref. 26) respond to glucocorticoid hormones because

* Present address: New York University Medical Center, Department of Environmental Medicine, New York, New York 10016, USA.

Fig. 2 The SV40 enhancer confers TPA responsiveness upon the heterologous TK gene promoter. *a*, Kinetic analysis of the induction response by RNA blot hybridization. HepG2 cells were transfected with a mixture of pSVE-/TK, a plasmid in which the SV40 enhancer is inserted in the antisense orientation upstream to the TK promoter³², pSV2-CAT and α -globin as described above. Twenty-four hours after the glycerol shock, TPA was added to one-half of the cultures and cells were collected for RNA extraction at the indicated time points. 50- μ g samples of total cellular RNA were analysed by Northern blot hybridization for expression of the transfected SVE-TK and α -globin genes and the endogenous α -tubulin gene. - and + signs indicate the absence or presence of TPA. As a control for the enhancer effect, the enhancerless TK gene was co-transfected with the other two plasmids and its expression was analysed after 8 h of induction, as described above. The positions of the 28S and 18S ribosomal RNAs are indicated. *b*, Primer extension analysis. To determine whether the TK mRNA was correctly initiated from the authentic HSV-TK promoter³³, the start site of transcription was determined by primer extension as described previously^{27,32}. Because we used a synthetic TK primer complementary to positions +58 to +81 of the TK gene, the correctly initiated transcript should yield an extension product of 81 bases, as was observed. *c*, Time-course of the induction response. The level of expression of TK mRNA in the absence or presence of TPA for each of the time points indicated was determined by densitometry, with a Zeineh soft-laser scanner, and normalized to the expression of the co-transfected α -globin gene to yield the relative fold of induction (TK/ α -globin). *d*, Effect of macromolecular synthesis inhibitors on induction of pSVE-TK by TPA. HepG2 cells were transfected with pSVE-/TK, pSV2-CAT, and α -globin, as described above, and 24 h after the glycerol shock, cycloheximide (300 μ g ml⁻¹) and actinomycin D (2 μ g ml⁻¹) were added to some of the cultures. Thirty minutes later TPA (100 ng ml⁻¹) was added, and after an additional 8 h incubation total cellular RNA was isolated and analysed by Northern blot hybridization with the TK, α -globin and α -tubulin probes, as described above. The expression of the SVE-/TK gene was normalized relative to that of the co-transfected α -globin gene to yield the relative levels of expression.



their control regions contain hormone-inducible enhancer elements. Similarly, metallothionein genes contain metal-inducible enhancer elements^{20,27,28}, and interferon genes contain enhancer elements that are activated by double-stranded RNA and viral infection²⁹⁻³¹. These results, and the observation that expression of pUCAT2, a plasmid in which an enhancerless SV40 early promoter controls expression of the CAT gene³², is not induced by TPA (see Tables 1 and 2) led us to examine the role of the SV40 enhancer in mediating the transcriptional response to TPA.

The results of two different sets of experiments indicate that the SV40 enhancer behaves like a TPA-inducible enhancer in HepG2 cells. First, we constructed four different derivatives of pUCAT2. These constructions harbour a single copy of the SV40 enhancer inserted in both orientations either upstream or downstream of the CAT transcription unit. As shown in Table 2, expression of each of the SV40 enhancer-containing constructs (pSV2-CAT, pO+/CAT, pO-/CAT, pCAT/O+ and pCAT/O-) is induced by TPA. Second, we linked the SV40 enhancer to the herpes simplex virus thymidine kinase (TK) gene to study the effect of the enhancer on a heterologous promoter. Two plasmids, pTK, containing the TK gene³³, and pSVE-/TK, containing the SV40 enhancer upstream to the TK promoter in the antisense orientation³², were transfected into HepG2 cells, and TK mRNA levels were determined by S1 nuclease protection. TPA treatment led to a 4.5-fold increase (determined by densitometry) in expression of correctly initiated TK mRNA from the enhancer-containing plasmid, pSVE-/TK. No effect of TPA on TK expression was observed using pTK (data not shown, but see Fig. 2), again demonstrating the specificity of the induction. We continued characterizing the induction response using pSVE-/TK. This plasmid was co-transfected into HepG2 cells with pSV2-CAT and α -globin. Twenty-four hours after transfection TPA was added and the time-course of induction was determined. As shown in Fig. 2a and c, expression of pSVE-/TK is rapidly stimulated by TPA. The level of TK mRNA peaks 2-4 h post-induction, then slowly declines to its basal level. The co-transfected α -globin gene also exhibits a modest response to TPA, but induction of pSVE-/TK is far

more extensive and is more than 5-fold higher than that of α -globin (Fig. 2c). The kinetics of pSV2-CAT mRNA induction is similar to that of pSVE-/TK (data not shown). The enhancerless plasmid, pTK, did not respond to TPA (Fig. 2a). Primer extension analysis indicates that the TK transcripts are correctly initiated from the authentic TK gene promoter both before and after induction (Fig. 2b).

To determine whether the response to TPA is a primary induction response, transfected HepG2 cells were incubated with an inhibitor of protein synthesis, cycloheximide (CHX) (30 μ g ml⁻¹), for 30 min before addition of TPA. CHX treatment alone led to induction of TK mRNA. However, addition of TPA to CHX-treated cells led to a further increase in TK mRNA, an effect that was additive with that of the inhibitor (Fig. 2a). This result shows that induction by TPA is independent of *de novo*

Table 1 Effect of DNA input on TAP induction of CAT expression

DNA input (μ g per plate)	CAT activity					
	pUCAT2			pSV2-CAT		
	Control	TPA	Times original level	Control	TPA	Times original level
0.5	0	0	—	53	1,022	19.4
2.0	0	0	—	733	4,917	6.7
5.0	1.1	1.8	1.6	3,331	17,351	5.2
10.0	8.6	9.8	1.1	6,027	22,572	3.7

pUCAT2 or pSV2CAT DNA in amounts indicated were transfected into HepG2 cells as described in the legend to Fig. 1. The total amount of transfected DNA was kept constant by the addition of pUC13 DNA. Cells were induced, collected and processed for determination of CAT activity as described above. Shown are the absolute values of CAT activity expressed in the transfected cells as CAT units per μ g protein, and induction by TPA (times original level). In pUCAT2 the CAT transcription unit is controlled by an SV40 early promoter lacking the enhancer³².

Table 2 Promoter specificity of the TPA response

Plasmid	Relative CAT activity		Induction (times original level)
	Control	TPA	
pSV2-CAT	1.0	4.7	4.7
phMT-II _A -CAT	0.34	0.28	0.82
phMT-I _A -CAT	0.11	0.17	1.5
pI ₁₀ -CAT	1.0	1.2	1.2
pA1b-CAT	0.008	0.014	1.7
pRSV-CAT	0.56	0.55	0.98
pMSV-CAT	0.067	0.039	0.58
pUCAT2	0.009	0.015	1.6
pO ⁺ /CAT	0.22	0.81	3.7
pO ⁻ /CAT	0.29	0.83	2.9
pCAT/O ⁺	0.38	1.89	4.97
pCAT/O ⁻	0.34	1.54	4.53

The different plasmids listed in the table were transfected at 5 µg per plate of HepG2 cells as described above. CAT activity is expressed relative to the level of uninduced pSV2-CAT expression. This experiment was repeated at least three times with essentially the same results: only constructs containing the SV40 enhancer were significantly induced. phMT-II_A-CAT is a fusion of the hMT-II_A promoter to the CAT transcription unit²⁰, phMT-I_A-CAT is constructed similarly and contains the promoter of the hMT-I_A gene⁴², pI₁₀-CAT contains the promoter of the calcium-inducible gene 3C5 (ref. 43). pA1b-CAT contains the promoter of the rat albumin gene⁴⁴. pRSV-CAT and pMSV-CAT contain the LTRs of Rous sarcoma virus and Moloney sarcoma virus, respectively^{45,46}. To construct pU⁺/CAT, pO⁻/CAT, pCAT/O⁺ and pCAT/O⁻, the *Pvu*II site at position 270 of SV40, in pO (ref. 17) was converted to a *Bgl*II site by addition of a synthetic linker (M.K., unpublished). A *Bgl*II-*Bam*HI fragment which contained a single 72-bp repeat and 20 additional bp of DNA was excised and subcloned in either orientation into the *Bgl*II site (5' to the CAT transcription unit) of pUCAT2³², to yield pO⁺/CAT and pO⁻/CAT or into the *Bam*HI (3' to the CAT transcription unit) of the same vector to yield pCAT/O⁺ and pCAT/O⁻. The + and - signs refer to the orientation of the repeat relative to the early promoter.

protein synthesis. In contrast, inhibition of RNA synthesis, achieved by incubation of transfected cells with actinomycin D (2 µg ml⁻¹), completely blocks the response.

The results presented demonstrate that, in HepG2 cells, the transcriptional stimulatory activity of the SV40 enhancer is increased as a specific response to treatment with the tumour promoter TPA. This effect is independent of *de novo* protein synthesis but requires ongoing RNA synthesis. Since the expression of pSV2-CAT also is decreased when the transfected cells are incubated in low serum medium, the enhancer also may be the target for the serum effect. While the response to serum was not analysed in detail, it is known from other systems that TPA can act as an indirect agonist of serum factors due to its ability to activate protein kinase C (ref. 14). Further experiments are required to determine whether activation of protein kinase C is an essential step in the pathway leading to transcriptional activation of the SV40 enhancer.

Recently Treisman³⁴ described a serum-responsive enhancer element in the 5'-flanking region of the *c-fos* gene, a gene whose transcription is increased after treatment of fibroblasts with serum, platelet-derived growth factor or TPA^{2,3}. A response to serum was also observed when the SV40 enhancer was placed in either orientation upstream or downstream from the *c-fos* gene, but the kinetics of the induction were different from those of the native *c-fos* gene. Since no distinction was made between the relative contributions of the *c-fos* promoter, 3'-flanking sequence (which has a role in *c-fos* mRNA destabilization), and the SV40 enhancer to the observed induction, it is not clear whether a common *trans*-acting factor is responsible for induction of both enhancers by serum or TPA. However, the striking difference in induction kinetics suggests that two different

mechanisms operate to regulate the highly transient response of the *c-fos* enhancer and the slower, more persistent response of the SV40 enhancer.

A possible mechanism to explain the induction of the SV40 enhancer in HepG2 cells by TPA could involve a post-translational modification of one or more of the factors that bind to the SV40 enhancer (ref. 35 and W. Lee *et al.*, in preparation). In the absence of TPA or serum the transcriptional stimulatory activity of this factor, present in low levels in HepG2 cells, is weak, but is increased after incubation with TPA, possibly due to activation of protein kinase C. The same factor could be present in a higher amount in TPA non-responsive cell lines and, even in the absence of TPA, this level is sufficient to confer full activation of the enhancer. However, determination of the exact mechanism of action will require the identification and isolation of this *trans*-acting factor. This task should be facilitated by the fact that the *cis*-element which serves as a target for the TPA effect, the SV40 enhancer, is extremely well characterized¹⁷⁻¹⁹ and some of the *trans*-acting factors which bind to it have recently been identified (ref. 35 and W. Lee *et al.*, in preparation).

TPA also is known to induce the expression of various DNA tumour viruses in latently infected cells³⁶. This induction has been demonstrated for Epstein-Barr virus^{37,38}, adenovirus³⁹, SV40 (ref. 40), and bovine papilloma virus⁴¹. The mechanism by which TPA and other tumour promoters induce the expression of these oncogenic viruses is not known, but it is possible that it involves activation of the enhancers in these viruses. In latently infected cells the level of viral gene expression is very low and insufficient to initiate either transformation or replication. Treatment of such cells with TPA may increase the activity of an enhancer element controlling early viral gene expression. This activation would lead to induction of early gene expression to a level permissive for either oncogenic transformation or initiation of viral replication. According to this model, tumour promoters can lead to malignant transformation not only by activating transcription of cellular oncogenes² but also by activating the enhancer elements of cryptic DNA tumour viruses.

We thank Drs Lou Laimins (pMSV-CAT and pRSV-CAT), Pierre Chambon (pO), Amy Lee (pI₁₀-CAT), Pamela Mellon (α-globin), Nick Cowen (α-tubulin) and Philippe Herbomel (pA1b-CAT) for plasmids and Drs Peter Herrlich and Peter Angel. This project was supported by grants from the NIEHS, the EPA and the Margaret Early Trust for Medical Research. M.K. is a Searle Scholar, and R.J.I. was supported in part by a PHS training grant to the Department of Microbiology, University of Southern California School of Medicine.

Received 10 March; accepted 8 August 1986.

1. Lau, L. F. & Nathans, D. *EMBO J.* **4**, 3145-3151 (1985).
2. Greenberg, M. E. & Ziff, E. B. *Nature* **311**, 433-438 (1984).
3. Cochran, B. H., Zullo, J., Verma, I. & Stiles, C. D. *Science* **226**, 1080-1082 (1984).
4. Cochran, B. H., Reffel, A. C. & Stiles, C. D. *Cell* **33**, 939-947 (1983).
5. Callahan, M., Cochran, B. H. & Stiles, C. D. in *Growth Factors in Biology and Medicine* (eds Evered, D., Nugent, J. & Whelan, J.) 87-97 (Pitman, London, 1985).
6. Scher, C. D., Shepard, R. C., Antoniadis, H. N. & Stiles, C. D. *Biochim. biophys. Acta* **560**, 217-241 (1979).
7. Pledger, W. J., Stiles, C. D., Antoniadis, H. N. & Scher, C. D. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4481-4485 (1977).
8. Persson, H., Gray, H. E. & Godeau, F. *Molec. cell. Biol.* **5**, 2903-2912 (1985).
9. Muller, R., Bravo, R., Burckhardt, J. & Curran, T. *Nature* **312**, 716-720 (1984).
10. Kruijer, W., Cooper, J. A., Hunter, T. & Verma, I. M. *Nature* **312**, 711-716 (1984).
11. Kelly, K., Cochran, B. H., Stiles, C. D. & Leder, P. *Cell* **35**, 603-610 (1983).
12. Stiles, C. D. *Cell* **33**, 653-655 (1983).
13. Brown, K. D. *et al. Biochem. biophys. Res. Commun.* **123**, 377-384 (1984).
14. Nishizuka, Y. *Nature* **308**, 693-698 (1984).
15. Berridge, M. J. *Biochem. J.* **220**, 345-360 (1984).
16. Habnicht, A. J. R. *et al. J. biol. Chem.* **256**, 12329-12335 (1981).
17. Zenke, M. *et al. EMBO J.* **5**, 387-397 (1986).
18. Khoury, G. & Gruss, P. *Cell* **33**, 313-314 (1983).
19. Herr, W. & Clarke, G. *Cell* **45**, 461-470 (1986).
20. Scholer, H. *et al. Science* **232**, 76-80 (1986).
21. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044-1051 (1982).
22. Graham, F. L. & van der Eb, A. J. *Virology* **52**, 456-467 (1973).
23. Castanaga, M. *et al. J. biol. Chem.* **257**, 7847-7851 (1982).
24. Chandler, V. L., Maler, B. A. & Yamamoto, K. R. *Cell* **33**, 489-499 (1983).
25. Ponta, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 1020-1024 (1985).

26. Karin, M. *et al.* *Cell* **36**, 371-379 (1984).
27. Karin, M. *et al.* *Molec. cell. Biol.* (submitted).
28. Serfling, E., Lubbe, A., Dorach-Hasler, K. & Schaffner, W. *EMBO J.* **4**, 3851-3859 (1985).
29. Ryals, J., Dierks, P., Ragg, H. & Weissmann, C. *Cell* **41**, 497-507 (1985).
30. Goodbourn, S., Zinn, K. & Maniatis, T. *Cell* **41**, 509-520 (1985).
31. Fujita, T., Ohno, S., Yasumitsu, H. & Taniguchi, T. *Cell* **41**, 489-496 (1985).
32. Haslinger, A. & Karin, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8572-8576 (1985).
33. McKnight, S. *Nucleic Acids Res.* **8**, 5949-5964 (1979).
34. Treisman, R. *Cell* **42**, 889-902 (1985).
35. Wildeman, A. G. *et al.*, *Molec. cell. Biol.* **6**, 2098-2105 (1986).
36. Zur-Hausen, H., Yamamoto, N. & Bauer, G. *Carcinogenesis* **7**, 617-623 (1982).
37. Zur-Hausen, H., O'Neill, F. J., Freese, U. K. & Hecker, E. *Nature* **272**, 373-375 (1978).
38. Eliasson, L., Nilsson, M. T., Kallin, B. & Klein, G. in *Cellular Interactions by Environmental Tumor Promoters* (eds Fujiki, H. *et al.*) 139-145 (Japan Sci. Soc. Press, Tokyo, 1984).
39. Fisher, P. B., Weinstein, I. B. & Ginsberg, H. S. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2311-2314 (1978).
40. Monier, R., Daya-Grosjean, L. & Sarasin, A. in *Carcinogenesis: Fundamental Mechanisms and Environmental Effects* (eds Pullman, B., Tso, P. & Gelboin, H.) 371-378 (Reidel, Dordrecht, 1980).
41. Amtmann, E. & Sauer, G. *Nature* **296**, 675-677 (1982).
42. Heguy, A., West, A., Richards, R. I. & Karin, M. *Molec. cell. Biol.* **6**, 2149-2157 (1986).
43. Resendez, E. Jr *et al.* *Molec. cell. Biol.* **5**, 1212-1219 (1985).
44. Ott, M. O. *et al.* *EMBO J.* **3**, 2505-2510 (1984).
45. Gorman, C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **79**, 6777-6781 (1982).
46. Laimins, L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **79**, 6453-6457 (1982).

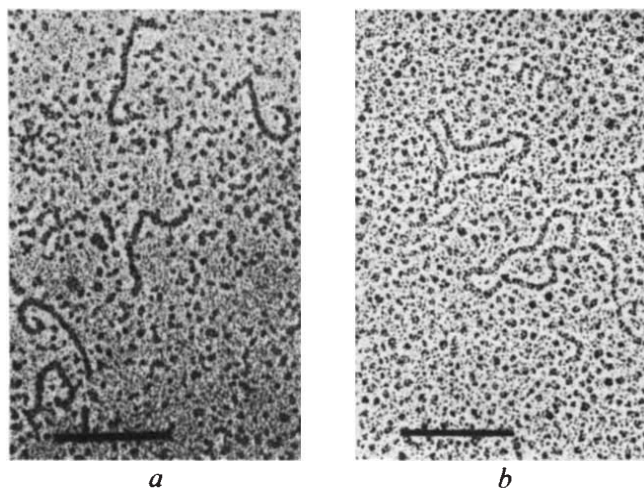


Fig. 1 Electron micrographs of HDV RNA. *a*, Non-denaturing; *b*, denaturing conditions. *a*, RNA was spread from 70% formamide, 0.12 M Tris-HCl (pH 8.0), 8 mM sodium EDTA, 0.01% cytochrome *c* on a hyperphase of triple distilled water. *b*, RNA was treated with 1 M glyoxal in 50% formamide, 0.1× SSC at 50 °C for 1 h and spread as in (*a*) after addition of 40 mM glyoxal. Bar, 0.2 μm.

The hepatitis delta (δ) virus possesses a circular RNA

A. Kos*, R. Dijkema†, A. C. Arnberg‡, P. H. van der Meide* & H. Schellekens*

* Department of Virology, Primate Center TNO, PO Box 5815, 2280 HV Rijswijk, The Netherlands

‡ Biochemical Laboratory, The University, Nijenborg 16, 9747 AG Groningen, The Netherlands

Hepatitis delta (δ) virus (HDV), a satellite virus of the hepatitis B virus (HBV), causes a severe form of viral hepatitis in humans¹. Here we present evidence based on electron microscopy and electrophoretic behaviour that HDV contains a single stranded circular RNA molecule. This is the first animal virus identified with a circular RNA genome. Circular RNAs have only been found in plant viruses. We have obtained a partial complementary DNA clone representing ~25% of the total genome of HDV. Analysis of this cDNA revealed similarity to two plant viruses that may explain the origin of the virus.

HDV superinfects patients with an acute or chronic hepatitis. It has been implicated in acute fulminant hepatitis B², and, more importantly, infects chronic carriers of hepatitis B^{3,4}. Up to 20-30% of all exacerbations of chronic carriers of HBV have been reported to be caused by hepatitis delta virus⁵.

HDV is a 36-nm particle containing a small RNA molecule of 1.75 kilobases (kb) (1/4 the size of the smallest RNA virus and 3-7 times larger than viroids) with a coat of hepatitis B surface antigen molecules from the helper HBV^{6,7}. An HDV antigen of *M_r* 68,000, initially identified⁸, appears to be two HDV-specific proteins of *M_r* 27,000 and 29,000⁹. From the sizes of the HDV proteins and the length of the encapsulated RNA molecule one can speculate that the HDV RNA might only encode the core antigen.

We have analysed the HDV RNA using electron microscopy. Under native conditions (Fig. 1*a*) rod shaped linear molecules of 833 ± 111 nucleotides (*n* = 82) were seen, whereas under strongly denaturing conditions¹⁰ circular single stranded molecules (Fig. 1*b*) with a contour length of 1,667 ± 154 nucleotides (*n* = 41) were observed. The native molecule is probably double stranded because of internal base-pairing.

When electrophoresed on a 0.6%-agarose gel (Fig. 2, left) native HDV RNA has the same mobility as a double-stranded fragment of 880 base pairs (bp). Under denaturing conditions (2.2 M formaldehyde, 1% agarose) the RNA shows a slightly

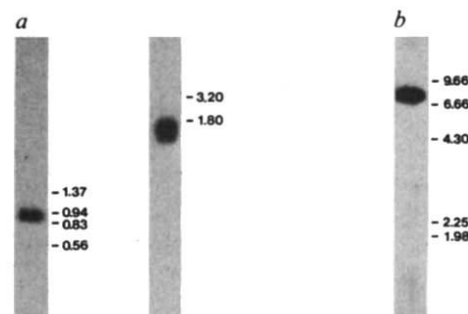


Fig. 2 Northern blot hybridization of nucleic acid samples with ³²P-labelled nucleotides. HDV RNA was size separated on 0.6% agarose (*a*, left), 1.0% agarose-2.2 M formaldehyde (*a*, right) or 4% polyacrylamide-7 M urea (*b*) gels and hybridized with an HDV cDNA clone (mD7) (Fig. 4, legend).

Methods. A 22-year-old male chimpanzee was acutely infected with HBV (ayw). Seventy days after inoculation the chimpanzee was superinfected with serum containing HDV particles from another chimpanzee. Three weeks after the HDV infection, serum samples were layered onto a 4-ml cushion of 20% (w/w) sucrose in 0.01 M phosphate, pH 7.9/0.85% NaCl and ultracentrifuged at 200,000g for 5 h at 4 °C in a Beckman SW41 rotor¹⁸. Pelleted particles were extracted for nucleic acid content by resuspension in 0.2 M NaCl 0.01 M EDTA, 2% SDS, 0.05 M Tris-HCl, pH 8.0, 1 mg ml⁻¹ proteinase K and incubated overnight at 37 °C. After extraction with an equal volume of phenol and chloroform, the aqueous phase was precipitated with 96% ethanol. The positions of λ markers and *Aspergillus* ribosomal RNA markers (*a*, right) are indicated in kilobases.

faster migration rate than 18S ribosomal RNA, indicating a size of 1,750 nucleotides (Fig. 2, middle lane; ref. 7). HDV RNA separated on polyacrylamide gels in the presence of urea behaved similarly to the circular RNAs of viroids¹¹: 4% polyacrylamide gel electrophoresis under denaturing conditions (7 M urea) reveals a migration rate comparable to an 8.2-kb fragment (Fig. 2, right) indicating a circular rather than linear molecule¹². The electrophoretic behaviour, together with the structure seen in the electron microscope, leads us to conclude that HDV is a circular single-stranded RNA virus.

RNA isolated from serum particles was used for cDNA synthesis¹³ using a nick in the circular structure of HDV RNA to

† Present address: Organon, PO Box 20, 5340 BH Oss, The Netherlands.

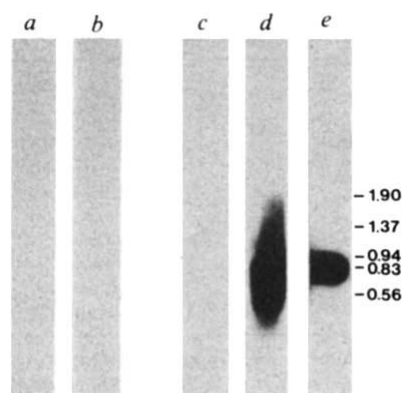


Fig. 3 Hybridization of pDelta-1 to DNA or RNA from infected chimpanzees. *a*, DNA, *b*, DNA, *c*, RNA from HBV infected chimpanzee; *d*, RNA and *e*, serum HDV RNA from HBV/HDV infected chimpanzee.

Methods. Chromosomal DNA from the liver of an HBV (*a*) or an HBV/HDV (*b*) infected chimpanzee was digested with *EcoRI* and separated on a 0.6% agarose gel in 1×TBE (90 mM Tris-HCl, 90 mM borate, 1.9 mM EDTA, pH 8.3). RNA from the liver of the HBV-infected chimpanzee (*c*) or the HBV/HDV-infected chimpanzee (*d*) and HDV from serum (*e*) of the HBV/HDV-infected chimpanzee was electrophoresed under the same conditions as the chromosomal DNA. All preparations were blotted and probed with the cDNA sequence from Delta-1.

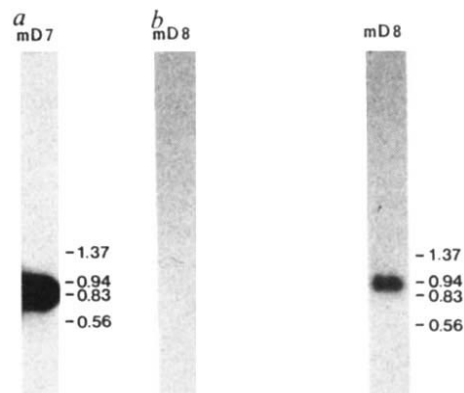


Fig. 5 Northern blot hybridization of HDV RNA with the strand-specific DNA probes mD7 (*a*, 2 h, -70°C exposure) and mD8 (*b*, 2 h, -70°C and *c*, 6 days -70°C exposure). Nonradioactive lambda *HindIII/EcoRI* markers are indicated in kilobases. HDV RNA was isolated from the serum as described in Fig. 2 and separated on a 0.6% agarose gel, 1×TBE (Fig. 3, legend).

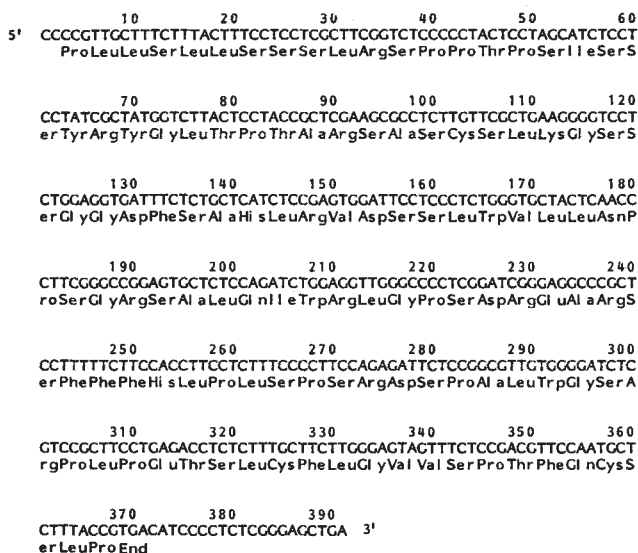


Fig. 4 Nucleotide sequence of the cDNA insert present in pDelta-1. The sequence was determined according to the Maxam-Gilbert method¹⁹. The largest predicted peptide is indicated. A *PstI* fragment containing the nucleotide sequence and its GC-tails was recloned into M13mp11 to give plasmids mD7 and mD8.

prime cDNA synthesis. The cDNA was cloned in pBR322 with dC-dG tails and propagated in *Escherichia coli* using standard techniques¹⁴. Differential hybridization with ³²P-labelled RNA derived from acute phase serum revealed one positive clone. This clone (pDelta-1) was used as a probe to screen a Southern blot of fractionated DNA from the liver of an HBV or HBV/HDV infected chimpanzee (Fig. 3). In lane *a* no hybridization was detected; the cDNA sequence of pDelta-1 is thus not normally present in the chimpanzee. We have no evidence for a cDNA intermediate of the hepatitis delta virus as there is no hybridization in lane *b*. The pDelta-1 was also used as a probe to screen a Northern blot of fractionated RNA from the livers of the same chimpanzees and RNA from the serum of the

HBV/HDV infected chimpanzee. Hybridization was only found with RNA isolated from the HBV/HDV infected chimpanzee (Fig. 3, lanes *d*, *e*). The smear in the RNA from the liver is probably caused by replication intermediates of HDV with different lengths and structures from the genomic RNA present in the virion (J. Taylor, personal communication).

The insert of pDelta-1 was estimated to be 520 bp, including dC-dG tails. The partial HDV cDNA sequence is presented in Fig. 4. The fragment has a G+C level of 57% and one hairpin with a stem of 19 bp with 6 mismatches and a loop of 65 nucleotides. Only one open reading frame gave a stretch of appreciable length (122 amino acids) with a stop codon (TGA) 70 bp before the end of the cDNA at nucleotide 370 (Fig. 4). The translation product of (part of) this hypothetical gene is given in Fig. 4. It contains three hydrophobic regions (amino acids 52–59, 79–88 and 105–113) and one strongly hydrophilic region located between amino acids 72–79. To establish that HDV RNA is single stranded and thereby identify the sequence equivalent to the HDV RNA we recloned the cDNA insert into bacteriophage M13mp11, obtaining two strand-specific DNA probes mD7 and mD8. Both M13-probes were hybridized with HDV RNA separated on 0.6% agarose gels. The mD7 probe hybridized strongly (Fig. 5) indicating that the native RNA has a sequence complementary to that shown in Fig. 4. Only after long exposure (6 days at -70°C) was a weak hybridization found between HDV RNA and mD8. This hybridization could be due to the packaging of antigenomic strands or to complementary sequences in the HDV RNA, which we know shows a strong secondary structure.

A circular RNA molecule can result in longer open reading frames than a linear RNA molecule of similar size but, assuming that the open reading frame in the cDNA clone represents the coding sequence of HDV RNA, these results indicate that hepatitis delta virus is a negative strand virus. This result is consistent with the inability of circular RNAs to attach to eukaryotic ribosomes¹⁵.

Comparison of the sequence of the cDNA clone pDelta-1 with sequences from Genbank (Los Alamos, United States) reveals considerable similarity to two sequences of the gemini virus cassava latent virus (CLV) DNA1 (strain West Kenyan 844). One of the sequences of CLV that is similar to the HDV cDNA is thought to be involved in DNA replication; the other is part of a sequence coding for a protein (M_r 15,000) of unknown function¹⁶. Homology was also found with the RNA plant virus tobacco mosaic virus (strain vulgare). This homology is restricted to a sequence of 17 bp, but a 166-bp cDNA sequence of HDV RNA (pKD3) was recently described¹² which contained

a different region from pDelta-1 and showed homology with sequences encoding the 30K and coat proteins of the cow pea strain of tobacco mosaic virus¹⁷.

These homologies with plant viruses and the viroid character of the genome of HDV encourages speculation about its possible plant origin. The question of the origin of this unique and intriguing virus may be answered by further study.

We thank Janny de Wit for electron microscopy. Supported by the Dutch Ministry of Economical Affairs, the Netherlands Foundation for Chemical Research (SON) and the Netherlands Organization for the Advancement of Pure Research (ZWO).

Received 19 June; accepted 29 August 1986.

- Rizetto, M. *et al.* *Gut* **18**, 997–1003 (1977).
- Rizetto, M., Purcell, R. H. & Gerin, J. L. *Lancet* **i**, 1215–1218 (1980).
- Rizetto, M. *et al.* *Lancet* **ii**, 986–990 (1979).
- Smedile, A. *et al.* *Gastroenterology* **81**, 992–997 (1981).
- Hadziyannis, S. J., Sherman, M., Lieberman, H. M. & Shafritz, P. A. *Hepatology* **5**, 544–547 (1985).
- Bonino, F. *et al.* *Hepatology* **1**, 127–131 (1981).
- Hoyer, B., Bonino, F. & Ponsetto, A. in *Viral Hepatitis and Delta Infection* (eds Verme, G., Bonino, F. & Rizetto, M.) 91–99 (Liss, New York, 1983).
- Rizetto, M., Shih, J.W.-K. & Gerin, J. L. *J. Immun.* **125**, 318–324 (1980).
- Bonino, F. *et al.* *J. Virol.* **58**, 945–950 (1986).
- Arnberg, A. C., Van Ommen, G.-J. B., Grivell, L. A., Van Bruggen, E. F. J. & Borst, P. *Cell* **19**, 313–319 (1980).
- Randles, J. W. in *Subviral Pathogens of Plants and Animals: Viroids and Prions* (eds Maramorosch, K. & McKelvey Jr, J. J.) 39–74 (Academic, Orlando, 1985).
- Denniston, K. J. *et al.* *Science* **232**, 873–875 (1986).
- Dijkema, R., Pouwels, P., Reus, A. de & Schellekens, H. *Nucleic Acids Res.* **12**, 1227–1242 (1984).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Kozak, M. *Nature* **280**, 82–85 (1979).
- Stanly, J. & Gay, M. R. *Nature* **301**, 260–262 (1983).
- Goelet, O. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **79**, 5818–5822 (1982).
- Rizetto, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **77**, 6124–6128 (1980).
- Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).

The nuclear lamina is a meshwork of intermediate-type filaments

Ueli Aebi*, Julie Cohn, Loren Buhle* & Larry Gerace

Department of Cell Biology and Anatomy, Johns Hopkins University School of Medicine, 725 North Wolfe St, Baltimore, Maryland 21205, USA

The nuclear lamina, a protein meshwork lining the nucleoplasmic surface of the inner nuclear membrane^{1,2}, is thought to provide a framework for organizing nuclear envelope structure³ and an anchoring site at the nuclear periphery for interphase chromatin^{3–5}. In several higher eukaryotic cells, the lamina appears to be a polymer comprised mainly of one to three immunologically related polypeptides of relative molecular mass (M_r) 60,000–75,000 (60–70K) termed lamins^{1,2}. Three lamins (A, B, and C) are typically present in mammalian somatic cells. Previous studies on nuclear envelopes of rat liver⁶ and *Xenopus* oocytes⁷ suggested that the lamina has a fibrillar or filamentous substructure. Interestingly, protein sequences recently deduced for human lamins A and C from complementary DNA clones^{8,9} indicate that both of these polypeptides contain a region of ~350 amino acids very similar in sequence to the coiled-coil α -helical rod domain that characterizes all intermediate-type filament (IF) proteins^{10,11}. Here we analyse the supramolecular organization of the native nuclear lamina and the structure and assembly properties of purified lamins, and show that the lamins constitute a previously unrecognized class of IF polypeptides.

We investigated the structure of the lamina by electron microscopy using manually isolated nuclear envelopes of *Xenopus* oocytes. After treatment of specimens with Triton X-100 to solubilize nuclear membrane components, the insoluble lamina with attached pore complexes is clearly visible in freeze-

Table 1 Biochemical properties of isolated rat liver lamins

	Lamins		
	A	C	B
Stokes radius (nm)*	10.6	10.2	11.1
s_{20} (S)†	4.6	4.6	4.9
	A and C		B
s_{20} (S)‡	4.5		ND
s_{20} (S)§	3.6		ND
M_r (K)	155.5 (dimer)		ND
	280.5 (tetramer)		
	216.0 (average)		

* Purified rat liver lamins A and C or lamin B (Fig. 2) were separately chromatographed at 4 °C on a 1.5 × 50 cm Sephacryl S-400 (Pharmacia) column in 0.2% Triton X-100, 20 mM Tris-HCl pH 9.0, 500 mM KCl, 2 mM EDTA, 1 mM dithiothreitol (DTT). Elution profiles were determined by quantitative analysis of SDS gels²⁴. Over a 10-fold range of different loading concentrations (resulting in lamin concentrations of 0.005–0.075 mg ml⁻¹ in peak fractions) each lamin eluted at an identical position as a roughly symmetrical peak, indicating that the lamins are dimers at these low concentrations (see below).

† Sedimentation coefficients measured by centrifugation of purified lamins A and C (0.75 mg ml⁻¹) or purified lamin B (0.5 mg ml⁻¹) on linear 5–20% sucrose gradients in column chromatography buffer (see above) according to the procedure of Martin and Ames²⁵ (conditions: 500 mM KCl, 0.2% Triton X-100, pH 9.0). The high protein concentrations maintained during gradient centrifugation favour occurrence of lamin tetramers.

‡§ Sedimentation velocity measurements made in a Beckman model E analytical ultracentrifuge using a 12-mm double-sector Kel-F cell. Samples of lamins A and C (0.15 mg ml⁻¹) were run at a speed of 52,000 r.p.m. at 20 °C in 25 mM Tris at 250 mM KCl, pH 8.5 (‡) and 500 mM KCl, pH 9.0 (§).

|| Low-speed sedimentation equilibrium (Beckman Model E analytical ultracentrifuge) was used to determine the M_r of lamins A and C. Solutions of protein concentration of 0.15 or 0.35 mg ml⁻¹ were analysed using 12- or 30-mm double-sector cells filled to a column height of 2–3 mm (conditions: 500 mM KCl, pH 9.0). At these protein concentrations the lamins are a mixture of dimers and tetramers, and samples were run at several speeds (8,000, 11,000 and 14,000 r.p.m.) to evaluate optimally the dimer, tetramer and average M_r . The percentage dimer and tetramer at equilibrium were determined from the approximate concentration of the lower M_r species (dimer) according to Yphantis²⁶. A partial specific volume of 0.73 ml g⁻¹ was assumed in all calculations. Original data for biochemical determinations provided on request.

dried/metal shadowed (Fig. 1a) and negatively stained (Fig. 1b) preparations. Interestingly, in well-preserved areas the lamina consists of two approximately orthogonal sets of 10.5 ± 1.5-nm-diameter filaments with an average crossover spacing of 52 nm. This quasi-regular meshwork of lamina filaments is particularly striking in places where pore complexes have been mechanically removed from the lamina (Fig. 1a, inset). However, the meshwork was very sensitive to mechanical forces during preparation and was extensively deformed in some regions of extracted nuclear envelopes (Fig. 1a, bottom). Washing Triton-treated specimens for several minutes in a low-ionic strength buffer before freeze-drying/metal shadowing (Fig. 1c) yielded a pronounced axial 'beading' every 25 ± 1 nm along the lamina filaments, probably related to an inherent structural repeat of the lamin polymer (see below).

Previous ultrastructural studies on isolated nuclear envelopes of *Xenopus* oocytes revealed 'pore-connecting' fibrils after detergent and other extraction procedures⁷, although a regular higher order arrangement of fibrils was not detected, presumably because of mechanical disruption of its fragile organization during sample preparation.

We isolated lamin B and a mixture of lamins A and C from nuclear envelopes of rat liver (Fig. 2) and used them to determine some biochemical properties of lamins (Table 1). The presence of Triton X-100 or urea (used during isolation steps) was not required to keep purified lamins A and C soluble, and we could

* Present addresses: Biozentrum, University of Basel, Basel CH-4056 Switzerland (U.A.); Pennsylvania Muscle Institute, University of Pennsylvania, Philadelphia 19104, USA (L.B.).

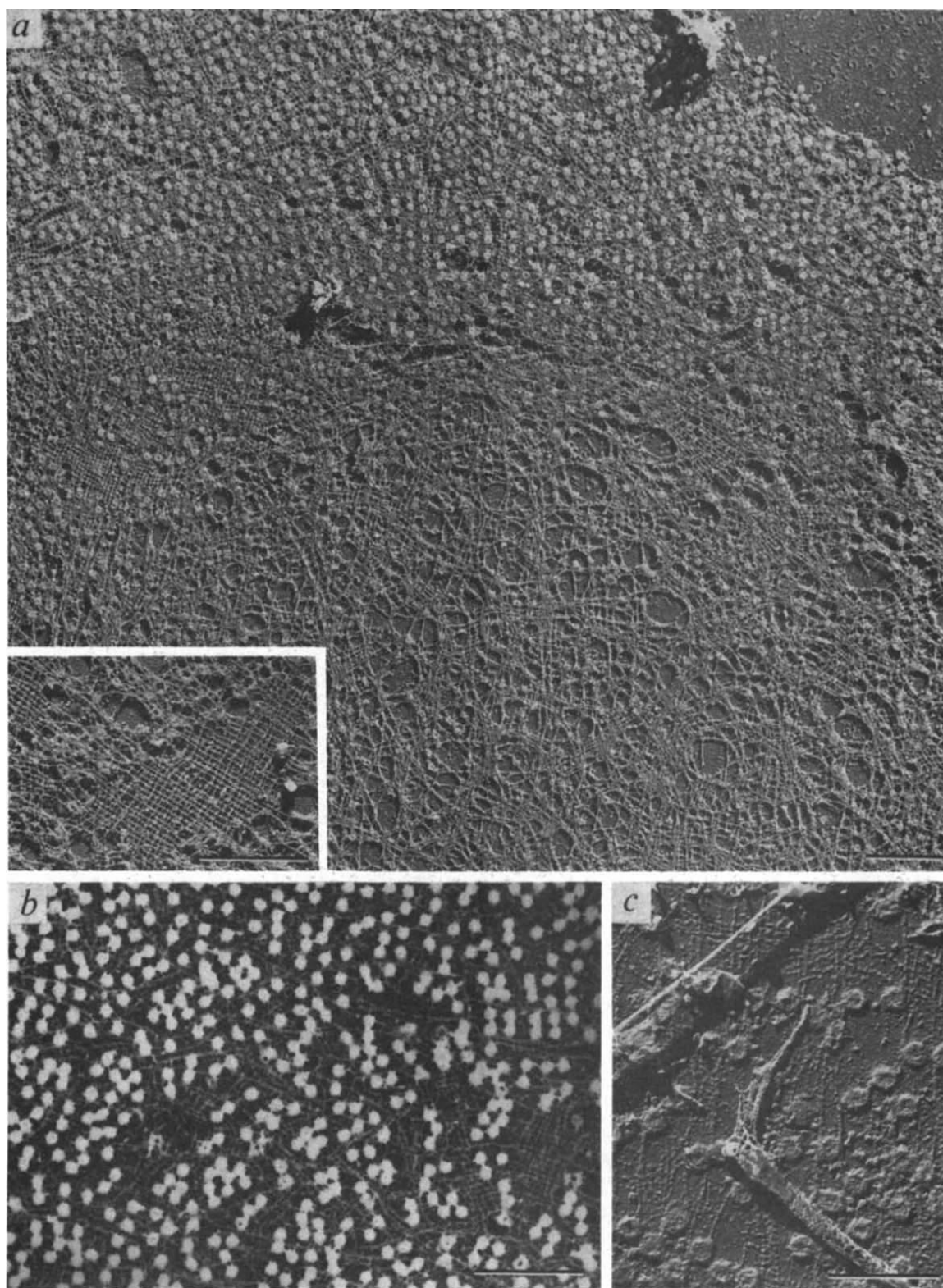


Fig. 1 Native nuclear lamina of *Xenopus* oocytes. *a*, Freeze-dried/metal-shadowed nuclear envelope extracted with Triton X-100, revealing the nuclear lamina meshwork partially covered with arrays of nuclear pore complexes (top). Inset, relatively well-preserved area of the meshwork from which pore complexes have been mechanically removed, comprised of two sets of near-orthogonal filaments. *b*, Same type of sample to *a* but negatively stained and imaged in dark-field mode. *c*, Similar preparation to *a* but extensively washed with low-salt buffer (LSB) before freeze-drying/metal shadowing, thereby revealing a 25-nm beading along the lamina filaments. Bars, 1 μm (*a* and inset, *b*), and 0.5 μm (*c*).

Methods. Mature oocytes were isolated from ovaries of *Xenopus laevis* and stored in Barth's saline solution²⁷ at 25 °C. Individual oocytes were put into LSB (10 mM HEPES, 1 mM KCl, 0.5 mM MgCl_2 , pH 7.5) and nuclei manually isolated²⁸ and cleaned of yolk and other debris. Nuclei were then transferred onto a carbon-coated grid containing 5–10 μl LSB. Once settled onto the support film, nuclei were opened with forceps and nuclear envelopes were gently spread across the grid with fine glass needles. Grids were then washed with 25 μl LSB and blotted with filter paper. To unveil the lamina, nuclear envelopes were extracted for 3–5 min with 0.1% Triton X-100 in LSB, washed for 30 s with 0.1% ethanol, and washed for either 30 s (*a*, *b*) or 4 min (*c*) with LSB. The material was fixed for 1 min with 1% glutaraldehyde followed for 1 min with 1% OsO_4 , both in LSB at 4 °C. After another 30-s wash at 4 °C with LSB, specimens were either freeze-dried/unidirectionally platinum/ carbon shadowed²⁹ or negatively stained with 0.75% uranyl formate³⁰. Magnification calibration according to Wrigley³¹.

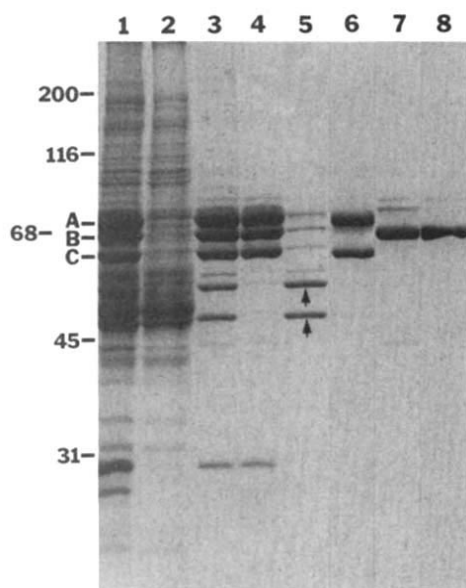


Fig. 2 Isolation of lamins from rat liver nuclear envelopes. All procedures at 0–4 °C unless indicated. Samples were electrophoresed on a 10% SDS-polyacrylamide gel and visualized by staining with Coomassie blue³. Nuclear envelopes were isolated and washed with 1 M KCl as described^{6,32}. Salt-washed nuclear envelopes (4–8 mg protein) (lane 1) were incubated for 30 min at a protein concentration of 0.25 mg ml⁻¹ in 10% sucrose, 2% Triton X-100, 20 mM MES-KOH pH 6.0, 300 mM KCl, 2 mM EDTA, 1 mM DTT and pelleted at 6,000g for 20 min, yielding a supernatant (lane 2) and a lamina-enriched pellet (lane 3). The pellet was resuspended at 0.3 mg ml⁻¹ in 2% Triton X-100, 20 mM Tris-HCl pH 9.0, 500 mM KCl, 2 mM EDTA, 1 mM DTT, and after 30 min was centrifuged at 200,000g for 40 min, yielding a supernatant of solubilized lamins (lane 4) and an insoluble pellet that contains contaminating cytochromes (lane 5, arrows). The solubilized lamin sample was then mixed with 24 vol. 6 M urea, 20 mM Tris-HCl, pH 9.0 and passed over a 0.5-ml phosphocellulose column (Whatman P11) followed by a serially connected 0.5-ml DEAE-cellulose column (Whatman DE-52). Subsequent elution of the phosphocellulose column with 6 M urea, 20 mM Tris-HCl pH 9.0, 500 mM KCl, 2 mM EDTA and 1 mM DTT (buffer U1) yielded a fraction containing an ~1:1 mixture of lamins A and C (lane 6). The DEAE column was first pre-washed (at room temperature) with 8.5 M urea, 10 mM MES, pH 6.5, 10 mM glycine, 0.5 mM EDTA and 1 mM DTT (buffer U2), and the column eluted with buffer U1 to yield a fraction enriched in lamin B (lane 7). Alternatively, a fraction containing more highly purified lamin B was obtained by eluting the DEAE column with a 25–75 mM KCl gradient in buffer U2 and pooling appropriate fractions (lane 8).

therefore characterize some properties of the two polypeptides in the absence of detergents and/or urea. However, as lamin B appears to be inherently less soluble than the other two lamins (see Fig. 3), we conducted all biochemical analyses of this protein in the presence of Triton X-100.

Sedimentation equilibrium experiments show that lamins A and C form dimers that can further associate into tetramers in a concentration-dependent fashion (Table 1). This dimer-tetramer equilibrium for isolated lamins A and C also became evident from measurements of sedimentation coefficients of these proteins in different buffers and at different protein concentrations, where values ranging from 3.6 to 4.6S were obtained (Table 1). However, in appropriate solutions at protein concentrations below 0.1–0.2 mg ml⁻¹, lamins A and C predominantly exist as dimers with Stokes radii of 10.2–10.6 nm (Table 1). Although we did not directly measure the M_r of lamin B by sedimentation equilibrium, the similarity of its hydrodynamic (Table 1) and structural (Fig. 3) properties to those of lamins A and C suggests that the basic oligomeric unit of lamin B is also a dimer. Consistent with our data, the M_r of the major

proteins of a pore complex/lamina fraction of bovine liver was calculated to be 111–142K, based on hydrodynamic measurements following solubilization with sodium deoxycholate¹².

Electron microscopy of glycerol-sprayed/rotary-shadowed replicas of purified lamins A and C and lamin B reveals rod-shaped molecules (Fig. 3). Lamins A and C, which are dimers under the conditions of this experiment, predominantly occur as a 52 ± 2 -nm rod with two globular heads at one end (Fig. 3a). Lamin B dimers consist of a 50 ± 2 -nm rod with one (less frequently two) globular head at one end (Fig. 3b), and are partially associated into higher order oligomers. The lamin dimers are structurally similar to the double-headed myosin II of *Acanthamoeba* except the lamin tail is shorter (although of the same width) and the heads are smaller (Fig. 3a, inset). Taken together, the results of our biochemical and electron microscope analyses indicate that under appropriate conditions all three rat liver lamins form parallel, unstaggered dimers with an ~50-nm rod domain attached to a pair of distinct globular end-domains. The globular heads on rat lamins are probably equivalent to the extensive carboxy-terminal non- α -helical end-domains detected on sequences of human lamins A and C^{8,9}.

When a mixture of purified rat liver lamins A and C was dialysed into buffers of near-physiological ionic strength and pH, they spontaneously assembled into long ~10-nm wide filaments (Fig. 4a). These 10-nm lamin filaments are structurally similar to the IFs that are assembled *in vitro* from the isolated 68K neurofilament polypeptide in similar conditions (Fig. 4a, inset). Furthermore, brief dialysis of these samples against lower ionic strength buffer shows beading with a 24.5 ± 0.6 -nm repeat along the axis of the lamin filaments (Fig. 4b), similar to the 25-nm beading in the shadowed native lamina filaments of *Xenopus* oocytes (Fig. 1c) and to the characteristic 21–23-nm axial repeat of IFs^{10,13}.

After longer dialysis of samples against lower ionic strength buffer, most material on the electron microscope grids appeared as thicker beaded filaments or cables (Fig. 4c) and subsequently, as paracrystalline arrays (Fig. 4d). A distinct stain-excluding pattern with a 24.4 ± 0.3 -nm axial repeat is seen in these cables and paracrystals, which presumably arises at least in part from lateral association of the 10-nm lamin filaments observed at earlier incubation times (see Fig. 4b). The stain-excluding regions along lamin filaments and paracrystals probably arise from superposition of globular carboxy-terminal domains of lamins, and the 24.5-nm axial repeat probably reflects a half-staggered arrangement of the ~50-nm-long lamin dimers in these structures. The paracrystalline arrays that are assembled from purified lamins A and C in our conditions appear similar to paracrystals generated from protein preparations of baby hamster kidney (BHK) cells enriched for lamins^{14,15}. However, in contrast to our results, the paracrystals of BHK lamins were reported to have an axial periodicity of 37–40 nm¹⁴ or 34–36 nm¹⁵.

The polypeptides comprising the different classes of IF all contain an ~310-amino-acid-long sequence of α -helical character flanked by non- α -helical end-domains of variable length^{10,11}. The α -helical domain forms a two-stranded parallel coiled-coil with the corresponding region of a neighbouring subunit, resulting in an unstaggered dimer with a 40–50-nm-long rod domain that is believed to form the backbone of all IFs. These dimers are thought to associate (possibly in an antiparallel fashion) into ~50-nm long tetramers which may in turn be arranged in a half-staggered fashion in the 10-nm IFs^{13,16}, with 4–10 tetrameric protofilaments per filament cross-section^{17,18}. Our results demonstrate that lamins possess all the major structural properties that have been described for IF polypeptides, at the levels both of protomers and of assembled filaments. Taken together with the extensive sequence similarity recently described^{8,9} for a region of human lamins A and C and the α -helical rod domain of IF polypeptides, our data indicate that lamins constitute IF proteins.

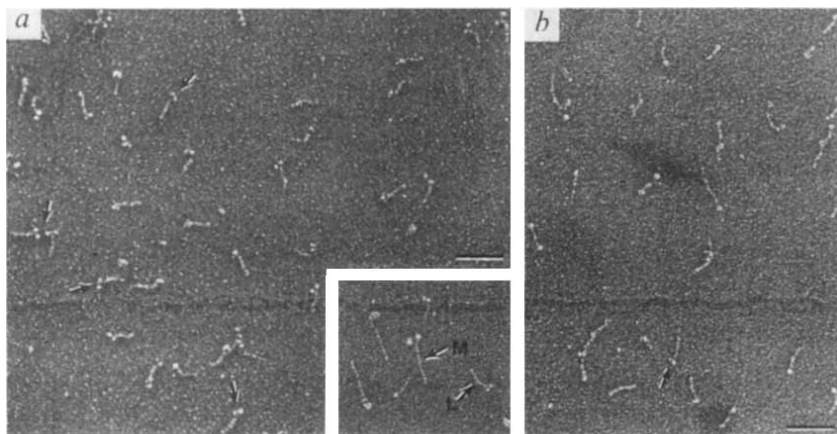
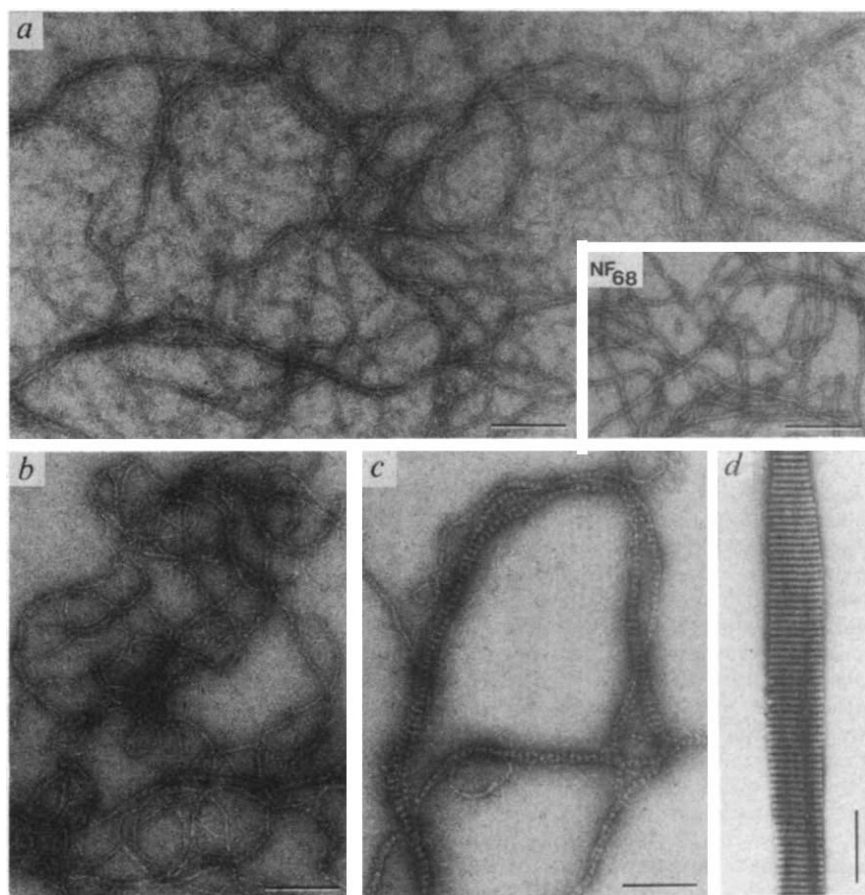


Fig. 3 Isolated rat liver lamins after glycerol spraying/rotary metal shadowing. *a*, Mixture (1:1) of isolated lamins A and C (see Fig. 2, lane 6) equilibrated in 25 mM Tris, 250 mM NaCl, 1 mM EGTA, 1 mM DTT, pH 8.5. Inset, 1:1 mixture (0.1:0.1 mg ml⁻¹) of lamins A and C and *Acanthamoeba* myosin II equilibrated in the same buffer as the lamins. The myosin (M) served as internal calibration standard (with an average tail length of 96 ± 2 nm³³) for the lamins (L). *b*, Isolated lamin B (see Fig. 2, lane 8) equilibrated in 25 mM Tris, 500 mM NaCl, 1 mM EGTA, 1 mM DTT, pH 8.5. Samples were prepared for electron microscopy by first diluting them to 0.05 mg ml⁻¹ with equilibration buffer and adding glycerol to a final concentration of 30% before they were sprayed onto freshly cleaved mica and further processed and imaged as described²⁹. Scale bars, 100 nm.

Fig. 4 *In vitro* reconstitution of IF and related structures from isolated rat liver lamins A and C. *a*, Lamins A and C (typically at a concentration of 0.25 mg ml⁻¹) were equilibrated in 25 mM Tris, 250 mM NaCl, 1 mM DTT, pH 8.0 at 37 °C and were then dialysed for ≥ 1 h at 37 °C against 25 mM MES, 200–250 mM NaCl, 1 mM DTT, pH 6.5. Inset, 10-nm neurofilaments reconstituted *in vitro* from the pure 68K neurofilament polypeptide isolated from bovine spinal cord³⁴ are shown for comparison. The lamin sample equilibrated at pH 6.5 as described above was dialysed by open dialysis at 37 °C against 25 mM MES, 150–175 mM NaCl, 1 mM DTT, pH 6.5 and aliquots removed at various time points: *b*, 15–30 min; *c*, 30–60 min; *d*, >60 min. Similar results to those in *b–d* were obtained when the material equilibrated at pH 8.0 (see above) was directly dialysed into 25 mM MES, 150–175 mM NaCl, 1 mM DTT, pH 6.5. The material shown in *a*, a mixture of IF and protofilamentous aggregates, was stable for up to 24 h. In contrast, in lower salt (150–175 mM NaCl) at pH 6.5 all the material eventually aggregated into huge paracrystalline arrays (*d*) visible as a macroscopic precipitate. Bars, 250 nm (*a–d*). Lamins A and C were isolated as in Fig. 2. For microscopy, 3–5- μ l aliquots were adsorbed to lightly glow-discharged and cytochrome *c* (0.2 mg ml⁻¹ in 0.1% 3-methyl-1-butanol) pre-treated carbon-coated grids, negatively stained with 0.75% uranyl formate and imaged as described³⁰.



The *Xenopus* lamina filaments in our preparations are probably comprised mainly of the single major M_r 66K lamin (L_{III})¹⁹ that is present in *Xenopus* oocytes. Lamin L_{III} is biochemically² and immunologically²⁰ related to somatic cell lamins, and is probably structurally similar to the latter as it is organized into filaments displaying an axial repeat (25 nm) nearly identical to the repeat obtained for structures assembled *in vitro* from rat liver lamins A and C.

Thin-section electron microscopy suggests that the lamina of rat hepatocytes is comprised of a filamentous meshwork similar to that of *Xenopus* oocytes (see Fig. 3c in ref. 6). Thus, it is possible that a meshwork of orthogonally oriented IFs is a common structural characteristic of the nuclear lamina. Although most cells (similar to rat hepatocytes and *Xenopus* oocytes) seem to contain an ~10-nm-thick lamina apposed to the inner nuclear membrane, some cell types contain an

unusually thick lamina (varying over 20–200 nm in diameter, see ref. 21). In these cases, the lamina may contain many layers of lamina filaments.

Of the three major lamins (A, B and C) typically present in mammalian somatic cells, lamin B has biochemical and immunological differences from the more closely homologous lamins A and C^{1,22}, and appears to be specialized for attaching the lamina to the inner nuclear membrane^{1,23}. Other functional properties of individual lamins have not yet been described. For functional reasons, we favour the possibility that lamin filaments are hybrid structures that contain all three lamins. Assuming that individual lamins carry out partially distinct functions, this arrangement could most easily explain how the lamina coordinately interacts with both chromatin and membranes.

We thank Mr E. C. Glavaris and Ms M. J. Meyer for technical assistance and Mr Ariel Lustig (Biozentrum, Basel) for perform-

ing the analytical ultracentrifugation experiments. We also thank Dr Tom Pollard for constructive comments on the manuscript. Supported by the National Institute for General Medical Sciences through grants GM31940 (U.A.) and GM28521 (L.G.), by a Searle Scholars Award (L.G.) and by a research award from the Maurice Muller Foundation of Switzerland (U.A.).

Received 1 April; accepted 29 July 1986.

1. Gerace, L., Comeau, C. & Benson, M. *J. Cell Sci. Suppl.* **1**, 137-160 (1984).
2. Krohne, G. & Benavente, R. *Expl. Cell Res.* **162**, 1-10 (1986).
3. Gerace, L., Blum, A. & Blobel, G. *J. Cell Biol.* **79**, 546-566 (1978).
4. Hancock, R. & Hughes, M. *Biol. Cell* **44**, 201-212 (1982).
5. Lebkowski, J. & Laemmli, U. *J. molec. Biol.* **156**, 325-344 (1982).
6. Dwyer, N. & Blobel, G. *J. Cell Biol.* **70**, 581-591 (1976).
7. Scheer, U., Kartenbeck, J., Trendelenburg, M., Stadler, J. & Franke, W. *J. Cell Biol.* **69**, 1-18 (1976).
8. McKeon, F., Kirschner, M. & Caput, D. *Nature* **319**, 463-468 (1986).
9. Fisher, D., Chaudhary, N. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
10. Weber, K. & Geisler, N. *Ann. N.Y. Acad. Sci.* **455**, 126-143 (1985).
11. Steinert, P., Steven, A. & Roop, D. *Cell* **42**, 411-419 (1985).
12. Havre, P. & Evans, D. *Biochemistry* **22**, 2852-2860 (1983).

13. Ip, W., Hartzer, M., Pang, S. & Robson, R. *J. molec. Biol.* **183**, 365-375 (1985).
14. Zachroff, R., Goldman, A., Jones, J., Steinert, P. & Goldman, R. *J. Cell Biol.* **98**, 1231-1237 (1984).
15. Goldman, A., Maul, G., Steinert, P., Yang, H. & Goldman, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3839-3843 (1986).
16. Geisler, N., Kaufmann, E. & Weber, K. *J. molec. Biol.* **182**, 173-177 (1985).
17. Eichner, R., Rew, P., Engle, A. & Aebi, U. *Ann. N.Y. Acad. Sci.* **455**, 381-402 (1985).
18. Steven, A., Trus, B., Hainfeld, J., Wall, J. & Steinert, P. *Ann. N.Y. Acad. Sci.* **455**, 371-379 (1985).
19. Krohne, G., Franke, W. & Scheer, U. *Expl. Cell Res.* **116**, 85-102 (1978).
20. Stick, R. & Hausen, P. *Cell* **41**, 191-200 (1985).
21. Fawcett, D. *Am. J. Anat.* **119**, 129-146 (1966).
22. Shelton, K., Higgins, L., Cochran, D., Ruffolo, J. & Egle, P. *J. biol. Chem.* **255**, 10978-10983 (1980).
23. Burke, B. & Gerace, L. *Cell* **44**, 639-652 (1986).
24. Fenner, C., Traut, R., Mason, D. & Wikman-Coffelt, J. *Analyt. Biochem.* **63**, 595-602 (1975).
25. Martin, R. & Ames, B. *J. biol. Chem.* **236**, 1372-1379 (1961).
26. Yphantis, D. *Biochemistry* **3**, 297-317 (1964).
27. Gurdon, J. *J. Embryol. exp. Morph.* **36**, 523-540 (1976).
28. Feldherr, C. & Richmond, P. *Meth. Cell Biol.* **17**, 75-79 (1978).
29. Fowler, W. & Aebi, U. *J. ultrastruct. Res.* **83**, 319-334 (1983).
30. Aebi, U., Fowler, W., Isenberg, G., Pollard, T. & Smith, P. *J. Cell Biol.* **91**, 340-351 (1981).
31. Wrigley, N. *J. ultrastruct. Res.* **24**, 454-464 (1968).
32. Gerace, L., Ottaviano, Y. & Kondor-Koch, C. *J. Cell Biol.* **95**, 826-837 (1982).
33. Pollard, T., Stafford, W. & Porter, M. *J. biol. Chem.* **253**, 4798-4808 (1978).
34. Geisler, N. & Weber, K. *J. molec. Biol.* **151**, 565-571 (1981).

***Rhizobium nod* genes are involved in inducing an early nodulin gene**

Francine Govers, Marja Moerman, J. Allan Downie^{*‡}, Paul Hooykaas[†], Henk J. Franssen, Jeanine Louwerse, Albert van Kammen & Ton Bisseling

Department of Molecular Biology, Agricultural University, De Dreijen 11, 6703 BC Wageningen, The Netherlands

^{*} John Innes Institute, Colney Lane, Norwich NR4 7UH, UK

[†] Department of Plant Molecular Biology, University of Leiden, The Netherlands

Rhizobium bacteria can invade the roots of leguminous plants and elicit the formation of root nodules. This process involves the expression of at least 20 nodule-specific genes encoded by the host plant, the so-called nodulin genes, which include the leghaemoglobin genes¹⁻³. During development of the root nodules the nodulin genes are differentially expressed¹. In fast-growing *Rhizobium* species several essential symbiotic genes are located on a large plasmid, and when fragments of the plasmid which contain the genes essential for nodulation (the *nod* region) are cloned and transferred to *Rhizobium* strains lacking the symbiotic plasmid ('cured' strains), the recipients regain the ability to form nodules. However, these nodules cannot fix nitrogen because essential nitrogen-fixation genes are absent⁴⁻⁶. We show here that transfer of the *nod* region (10 kilobases; kb) alone from a *Rhizobium leguminosarum* symbiotic plasmid into cured *Rhizobium* strains is sufficient to elicit nodulin gene expression in the host *Pisum sativum*. We also show that pea root nodules induced by an *Agrobacterium* strain containing the *R. leguminosarum* symbiotic plasmid express an early nodulin gene but no other nodulin genes. These results show that at least two signals are involved in the induction of expression of nodulin genes and the presence of the *Rhizobium* nodulation genes seems to be required to elicit the first signal.

A 60-kb region of DNA from the *R. leguminosarum* symbiotic plasmid pRL1J1 carries a cluster of nodulation genes which is flanked by two groups of genes involved in nitrogen fixation⁷. Downie *et al.*⁴ found that two overlapping cosmid clones of pRL1J1—pIJ1089 and pIJ1085—containing a 10-kb region in common, enabled a cured *R. phaseoli* strain to form nodules on pea; the development of these nodules appeared normal and bacteroid forms surrounded by a peribacteroid membrane were present, although the nodules did not fix nitrogen because of

nitrogen-fixation genes were absent. Transfer of the same cosmid clones to a cured *R. leguminosarum* strain also restored the pea nodulation ability of this strain. If the 10-kb region was absent no nodules were formed and it was concluded that the 10-kb region is the only fragment of the pRL1J1 symbiotic plasmid which is essential for pea nodule formation. We therefore used the strains containing pIJ1089 or pIJ1085 to determine whether pea nodulin genes are induced. Nodulin gene expression was studied in two ways, by Northern blot analyses and separation of *in vitro* translation products from nodule RNA on two-dimensional gels. We have shown previously that *in vitro* translation of pea nodule RNA gives rise to at least 21 nodulins¹. Figure 1a shows a two-dimensional gel of *in vitro* translation products of nodule RNA while Fig. 1b shows six major nodulin spots, including two leghaemoglobin (Lb) spots. Nodulin N-40' was detected on the gels 5 days before Lb whereas nodulins N-40 and N-68 appeared during nodule development simultaneously with the Lb spots. N-21 was only detectable 2 days later.

For the Northern blot analyses nodulin complementary DNA clones were selected representing either messenger RNAs that appear at the same day as Lb mRNA (pPsNod clones; F.G., unpublished results) or mRNAs that appear several days before Lb mRNA (pENOD clones, H.J.F., manuscript in preparation). The pPsNod clones are of pea origin whereas the pENOD clones were isolated from a soybean nodule cDNA library; however, one of these, pENOD2, cross-hybridizes strongly with RNA from pea nodules and represents a pea nodulin gene that is expressed 5 days before the Lb genes (H.J.F., manuscript in preparation), that is, at the same time as the *in vitro* translation product N-40' is first observed¹. Nodulin genes can thus be divided into two classes: the early nodulin genes, such as N-40' and ENOD2, and the nodulin genes that are expressed just before or after the onset of nitrogen fixation, such as N-68, N-40, N-21, Lb and PsNod genes.

In 21-day-old pea root nodules formed by a cured *R. phaseoli* strain containing pIJ1089 or pIJ1085 and no other parts of a symbiotic plasmid, the pattern of major (Fig. 1) and minor nodulin spots (data not shown) was similar on two-dimensional gels to that produced by the wild-type *R. leguminosarum* strain. N-68, N-40', N-40 and Lbs were present at wild-type levels, but the amount of N-21 was reduced in the nodules formed by the *Rhizobium* containing pIJ1085 or pIJ1089. On Northern blots, RNAs hybridizing to pENOD2 and the pPsNod clones were also detected (data not shown), and thus all the identified nodulin genes in pea are expressed. This shows that, apart from the *nod* region, *nif*, *fix* and other putative genes encoded by the symbiotic plasmid are not essential for the induction of expression of these nodulin genes. In previous experiments using *Rhizobium* mutants that produce ineffective (*fix*⁻) root nodules,

[‡] Present address: CSIRO, Division of Plant Industry, Canberra, Australia.

Fig. 1 Expression of nodulin genes in pea root nodules. Fluorographs of two-dimensional gels of *in vitro* translation products from total RNA isolated from: *a*, effective pea root nodules 17 days after sowing and inoculation with *R. leguminosarum* (PRE); *b*, uninfected pea roots from 8-day-old plants, pea root nodules induced by wild-type *R. leguminosarum* (PRE) (15 days), *R. phaseoli* cured of its own symbiotic plasmid and containing the cosmid pIJ1089 (21 days), and an *Agrobacterium* strain, LBA 2712, carrying a symbiotic plasmid (28 days). Nodules formed by LBA 2712 were first detected on roots of 21-day-old pea plants and were collected from 21-, 28-, 35- and 42-day-old plants, respectively. In none of these nodules were major nodulins detected by *in vitro* translation and in all cases bacteria isolated from surface-sterilized nodules were resistant to the appropriate antibiotics. The major nodulin spots N-68, N-40, N-40', N-21, Lb-1 and Lb-2 are indicated by arrowheads. In *b* only the parts of the gels within the squares indicated in *a* are shown. The nodulin pattern induced by *Rhizobium*/pIJ1085 (result not shown) is similar to that elicited by *Rhizobium*/pIJ1085. In *a*, ^{14}C -methylated M_r markers are shown on the left ($\times 10^{-3}$).

Methods. Nodulated pea plants (*Pisum sativum* var. Rondo) were grown in Leonard jars filled with Perlite under the growth conditions described by Bisseling *et al.*¹⁴. Total RNA was isolated from uninfected roots and nodules, and *in vitro* translated in the presence of ^{35}S -methionine in a mRNA-dependent rabbit reticulocyte lysate as described previously¹. The *in vitro* labelled translation products were separated by two-dimensional gel electrophoresis according to O'Farrell¹⁵, and 3-[(3-chloroamidopropyl)dimethylammonio]-1-propanesulphonate was used in the isoelectric focusing sample buffer¹⁶.

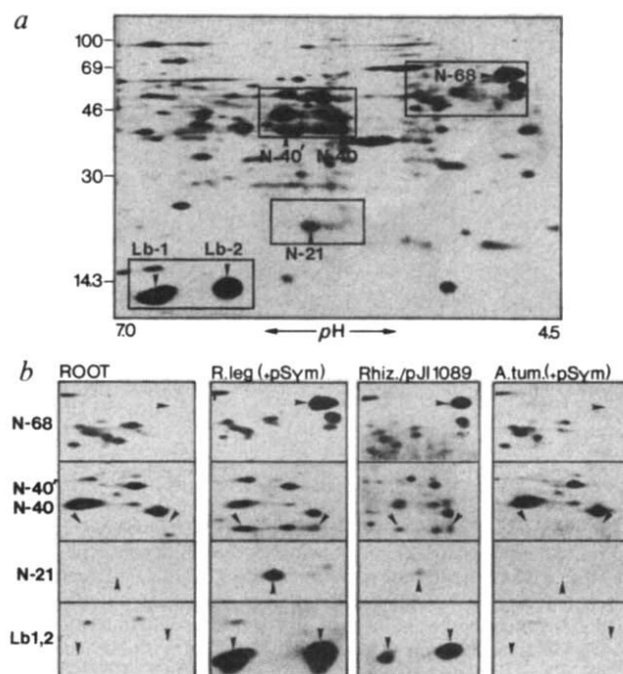
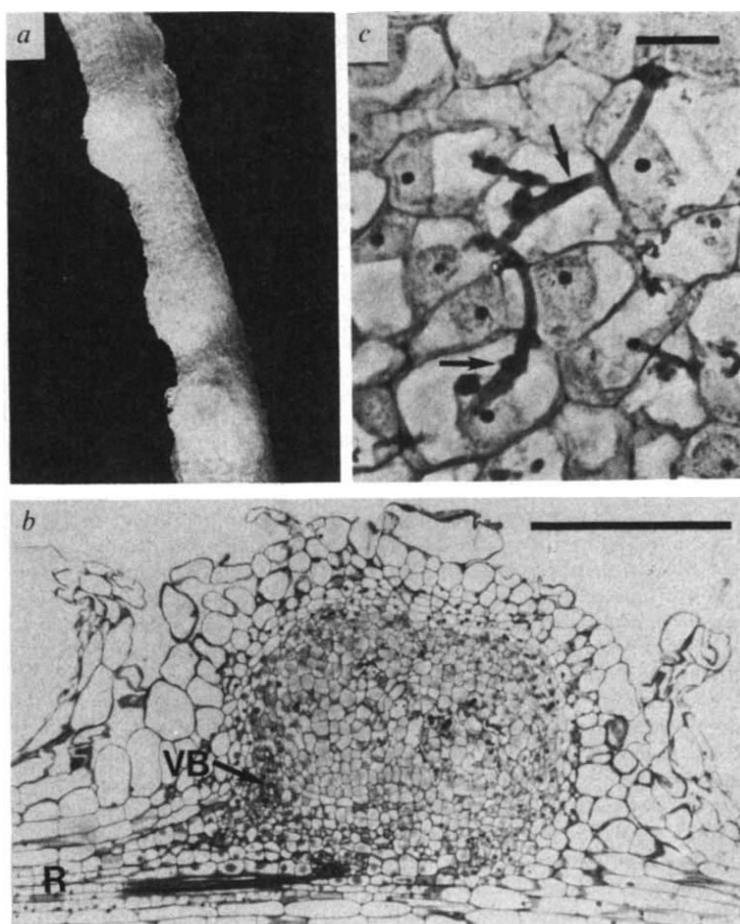


Fig. 2 Morphology and cytology of 'empty' nodules on pea roots induced by the *Agrobacterium* transconjugant LBA 2712. *a*, Part of a pea root with nodules (28 days). *b*, Light micrograph of a longitudinal section through a nodule (28 days). The vascular bundle (VB) connection to the main root (R) is maintained. Scale bar, 400 μm . *c*, Light micrograph of a part of the central nodule tissue in which infection thread-like structures are visible (arrows). Scale bar, 40 μm .

Methods. Pea plants were grown as described in Fig. 1 legend. Four-week-old nodules were fixed in 3% glutaraldehyde for 2 h, dehydrated in a graded ethanol series and embedded in Technovit 7100; 4- μm sections were cut with steel knives, placed on slides and stained with toluidine blue¹⁷.



nodulin N-21 was never found among the *in vitro* translation products and it was suggested that the expression of the N-21 gene was controlled by the nitrogen-fixation process¹. However, the present results refute this conclusion. The difference is probably due to the plant growth conditions. By growing the plants in Leonard jars containing perlite instead of trays containing gravel, we can detect N-21 at low levels.

The *nod* region is clearly essential for nodule development and, as shown above, the presence in *Rhizobium* of only this part of the *R. leguminosarum* symbiotic plasmid is sufficient to induce nodulin gene expression in pea. It is becoming increasingly evident that *Rhizobium* chromosome and non-symbiotic plasmids contain genetic information that is essential for effective nodule development. For example, mutations in the

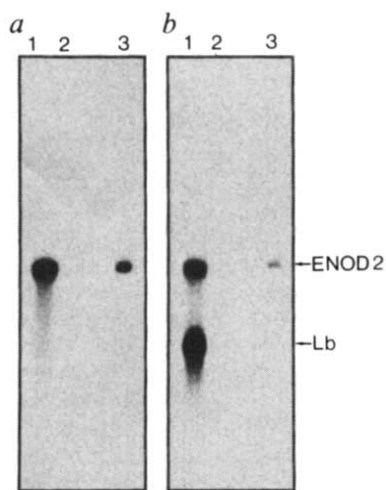


Fig. 3 Expression of leghaemoglobin and ENOD2 genes in pea root nodules. Autoradiographs of a Northern blot containing RNA from wild-type pea root nodules (17 days, lanes 1) induced by *R. leguminosarum* (PRE), 8-day-old uninfected pea roots (lanes 2) and 'empty' nodules induced by the *Agrobacterium* transconjugant LBA 2712 (28 days; lanes 3). The blot was hybridized with pENOD2 (a) and subsequently with pPsLb101 (b). ENOD2 RNA corresponds to 1,400 bases in length. Because the blot was not deprobed, the hybridization of the heterologous soybean cDNA clone pENOD2 with pea RNA shown in a is still visible in b but the intensity is reduced due to higher stringency washings after hybridization with the homologous Lb cDNA clone.

Methods. Samples of RNA (10 µg) were denatured in 0.5 M glyoxal, 50% dimethyl sulphoxide, 0.01 M sodium phosphate pH 7.0 for 60 min at 50 °C and separated on a 1.2% agarose gel in 10 mM sodium phosphate pH 7.0. After electrophoresis, the RNA was transferred to GeneScreen paper (NEN) using 25 mM sodium phosphate pH 6.5 as blotting buffer. The filter was prehybridized for 8 h in HB (50% formamide, 10× Denhardt's solution, 50 mM Tris-HCl pH 7.5, 1 M NaCl, 0.1% SDS, 10% dextran sulphate and 100 µg ml⁻¹ denatured salmon sperm DNA) and hybridized for 24 h in HB plus denatured pENOD2 DNA labelled with ³²P by nick translation¹⁸. The hybridization temperature was 35 °C and the filter was washed twice for 30 min at 42 °C in 2× SSC, 0.5% SDS and for 30 min at room temperature in 0.5× SSC, 0.1% SDS. After exposure to Kodak XAR-5 film the filter was prehybridized for 8 h, hybridized for 24 h at 42 °C with ³²P-labelled pPsLb101, washed twice for 30 min at 65 °C in 2× SSC, 0.5% SDS followed by 30 min at room temperature in 0.5× SSC, 0.1% SDS, and again exposed to Kodak XAR-5 film.

genes involved in the exopolysaccharide synthesis in *Rhizobium meliloti* result in nodules lacking intracellular bacteria⁸. Because it is not known whether the chromosome is involved in the induction of nodulin genes, we studied nodulin gene expression in pea root nodules formed by a strain (LBA 2712) that contains an *Agrobacterium* chromosome and a *R. leguminosarum* symbiotic plasmid (pSym1)⁹. It has been reported recently that 'empty' root nodules are formed on alfalfa roots by *Agrobacterium tumefaciens* carrying clones containing the *nod* region of *R. meliloti*^{5,10-12}, and we observed similar empty nodules on pea roots inoculated with LBA 2712. Histological examination of these small ineffective nodules showed that the plant cells lack bacteria, although some structures similar to infection threads were observed (Fig. 2). Vascular bundles were present at the periphery of the nodule, characteristic of normal nodules. These empty nodules first appeared 3 weeks after inoculation, and analyses of *in vitro* translation products from RNA isolated from empty nodules on 3-, 3.5-, 4- and 5-week-old plants revealed no Lb or other major nodulins (N-68, N-40, N-40' and N-21) (Fig. 1b). Hybridization of a Northern blot with pPsLb101, a pea Lb cDNA clone¹³, showed that Lb mRNA is found only in the wild-type nodules and not in the empty nodules (Fig. 3b), confirming the results found by *in vitro* translation of

the RNAs. However, hybridization of the same blot with pENOD2 showed that ENOD2 mRNA, an early nodulin mRNA, is present in 3.5-week-old empty nodules (Fig. 3a). Indeed, close examination of several two-dimensional gels revealed that a minor nodulin of the same relative molecular mass (M_r) as the hybrid-released translation product of pENOD2 (M_r 80,000; F.G., unpublished results) is consistently present in empty nodules. Hence, in empty pea root nodules only the early nodulin gene ENOD2 is expressed while the other nodulin genes, including that encoding the early nodulin N-40', are not activated. In pea tumours induced by *A. tumefaciens* no ENOD2, Lb or other nodulin mRNAs are detected (data not shown). In such tumours, the agrobacteria do not penetrate the plant cells and wounding is necessary for causing tumour development, so that the bacteria can reach the intercellular spaces. Furthermore, empty nodules contain no intracellular bacteria, although some infection threads with bacteria are observed (Fig. 2c). As the *Agrobacterium* strain carrying the symbiotic plasmid induces expression of the ENOD2 gene whereas *A. tumefaciens* does not, it seems unlikely that the *Agrobacterium* genome itself harbours genes involved in the induction of this gene. Such genes must be located on the symbiotic plasmid and, as shown by analyses of nodules induced by *Rhizobium* strains carrying pIJ1089 or pIJ1085, only a 10-kb region from the symbiotic plasmid is required for expression of the ENOD2 gene. This region contains the *nod* genes and we suggest that the *nod* gene products are responsible, directly or indirectly, for induction of ENOD2 gene expression.

We conclude from the present results that: (1) An intracellular location of *Rhizobium* is not required for the expression of the ENOD2 gene in pea, although this may be essential for expression of the other nodulin genes. (2) At least two signals from *Rhizobium* seem to be involved in expression of pea nodulin genes, one inducing the expression of the early nodulin gene ENOD2 and a second for the other nodulin genes; the generation of the latter signal might be dependent on the effects produced by the first signal. (3) The nodulation genes of *R. leguminosarum* are involved in the induction of the ENOD2 gene in pea. (4) *Rhizobium* genes not located on the symbiotic plasmid seem to be essential for the expression of other pea nodulin genes.

This work was supported financially by grants from the Netherlands Organization for the Advancement of Pure Science (M.M.), the Netherlands Organization for Biological Research (F.G.) and an Agrigenetics Sponsored Research Program (H.J.F. and J.L.). We thank A. Houwers and P. de Kam for culturing nodulated plants, G. Heitkönig, A. Jonkers and M. J. van Neerven for typing the manuscript and R. de Fluiter and P. Madern for preparing the figures.

Received 2 April; accepted 7 August 1986.

- Govers, F., Gloudemans, T., Moerman, M., Van Kammen, A. & Bisseling, T. *EMBO J.* **4**, 861-867 (1985).
- Fuller, F., Kunstner, P. W., Nguyen, T. & Verma, D. P. S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2594-2598 (1983).
- Bisseling, T., Been, C., Klugkist, J., Van Kammen, A. & Nadler, K. *EMBO J.* **2**, 961-966 (1983).
- Downie, J. A. *et al. Molec. gen. Genet.* **190**, 359-365 (1983).
- Hirsch, A. M. *et al. J. bact.* **158**, 1133-1143 (1984).
- Schofield, P. R., Ridge, R. W., Rolfe, B. G., Shine, J. & Watson, J. M. *Pl. molec. Biol.* **3**, 3-11 (1984).
- Downie, J. A., Ma, Q.-S., Knight, C. D., Hombrecht, G. & Johnston, A. W. B. *EMBO J.* **2**, 947-952 (1983).
- Finan, T. M. *et al. Cell* **40**, 869-877 (1985).
- Hooykaas, P. J. J., Snijderwint, F. G. M. & Schilperoord, R. A. *Plasmid* **8**, 73-82 (1982).
- Wong, C. H., Pankhurst, C. E., Kondorosi, A. & Broughton, W. J. *J. Cell Biol.* **97**, 787-794 (1983).
- Truchet, G. *et al. J. Bact.* **157**, 134-142 (1984).
- Hirsch, A. M., Drake, D., Jacobs, T. W. & Long, S. R. *J. Bact.* **161**, 223-230 (1985).
- Bisseling, T. *et al. in Advances in Nitrogen Fixation Research* (eds Veege, C. & Newton, W. E.) 579-586 (Nijhoff-Junk, The Hague, 1984).
- Bisseling, T., Van den Bos, R. C. & Van Kammen, A. *Biochim. biophys. Acta* **539**, 1-11 (1978).
- O'Farrell, P. H. *J. biol. Chem.* **250**, 4007-4021 (1975).
- Perdue, G. H., Schaub, H. W. & Selinonchick, D. P. *Analyt. Biochem.* **135**, 453-455 (1983).
- Gerrits, P. G. & Smid, L. *J. Microsc.* **132**, 81-85 (1983).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).